



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 116820

TO: Vanessa L Ford
Location: REM-3B25&3C18
Art Unit: 1645
Thursday, March 18, 2004

Case Serial Number: 09/543407

From: Barb O'Bryen
Location: Biotech-Chem Library
Remsen E01A69
Phone: 571-272-2518 *BOB*

barbara.obryen@uspto.gov

Search Notes

RUSH

THIS PAGE BLANK (USPTO)

STIC-Biotech/ChemLib

116820

From: Page, Thurman
Sent: Friday, March 12, 2004 2:35 PM
To: STIC-Biotech/ChemLib
Cc: Ford, Vanessa
Subject: FW: In re: 09/543407 sequence search

Importance: High

CRFE

Rush search approved.

Vanessa,
please resend the reason for rush to me for my record.
thanks

Thurman K. Page
SPE Art Units 1615 & 1616
Technology Center 1600

-----Original Message-----

From: Ford, Vanessa
Sent: Friday, March 12, 2004 1:52 PM
T : Page, Thurman
Subject: FW: In re: 09/543407 sequence search

Christina is out of the Office today.

-----Original Message-----

From: Ford, Vanessa
Sent: Friday, March 12, 2004 1:50 PM
To: Chan, Christina
Subject: In re: 09/543407 sequence search

Please search: SEQ ID NOs: 1, 2, 3, 4, 9, 11, 13, 15, 17, 19. please include interference searches. Please Rush.

Biotechnology Patent Examiner
Office: REM 3B25
Mailbox: REM 3C18
Phone: 571.272.0857
Art Unit: 1645

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-1

Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sls:*
12: gb_sts:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_ov:*
22: em_or:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_scs:*
28: em_un:*
29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_other:*
41: em_hcg_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	144909	144909 Sequence 58
2	456	100.0	2067	1 SEU43280	U43280 Salmone11
3	448	98.2	5103	1 STAJ2301	AJ002301 Salmone11
4	448	98.2	22411	1 AE008749	AE008749 Salmone11
5	446.4	97.9	254050	1 AL627269	AL627269 Salmone11
6	446.4	97.9	301983	1 AE016840	AE016840 Salmone11
7	440.6	96.6	1048	1 STAF8A	AU000514 Salmone11
8	331.8	72.8	361	6 144908	144908 Sequence 56
9	310.4	68.1	2889	1 CSP515700	AJ515700 Citrobact
10	262.8	57.6	2920	1 CFS151701	AJ515701 Citrobact
11	256	56.1	4680	1 ECCSGABDG	X90754 E.coli csgG
12	256	56.1	10346	1 AE000205	AE000205 Escherich
13	256	56.1	15047	1 D90741	D90741 Escherichia
14	254.4	55.8	456	6 AX814811	AX814811 Sequence
15	254.4	55.8	648	1 ECOC8GAA	ECOC8GAA Escherichia
16	247.8	54.3	1711	1 AF275733	AF275733 Escherichia
17	247.8	54.3	10190	1 AE005315	AE005315 Escherichia
18	247.8	54.3	306358	1 AE016759	AE016759 Escherichia
19	247.8	54.3	327773	1 AP002554	AP002554 Escherichia
20	229	50.2	230	1 SEU53207	U53207 Salmone11
21	217.6	47.7	2883	1 ESAS15702	AJ515702 Enterobac
22	190.4	41.8	10370	1 AE015131	AE015131 Shigella
23	190.4	36.7	437	1 AE016981	AE016981 Shigella
24	167.4	36.7	437	1 AF237726	AF237726 Shigella
25	93.2	20.4	19201	1 D90742	D90742 Escherichia
26	64.6	14.2	1212	1 AE0131756	AJ131756 Escherichia
27	48.2	10.6	78	6 AX814809	AX814809 Sequence
28	44.8	9.8	72	6 AX814798	AX814798 Sequence
29	44.4	9.6	2000	6 AX655393	AX655393 Sequence
30	36.2	7.9	1425	1 AF343445	AF343445 Lactobaci
31	36.2	7.9	302156	3 AC116977	AC116977 Dictyoste
32	36	7.9	168267	2 AC146009	AC146009 Pan trogl
33	35.6	7.6	301214	1 AE016786	AE016786 Pseudomon
34	34.6	7.6	11160	1 AE008209	AE008209 Agrobacte
35	34.6	7.6	11476	1 AE009405	AE009405 Agrobacte
36	34.4	7.5	151828	2 AC146008	AC146008 Pan trogl
37	34.2	7.5	1458	1 AF173043	AF173043 Oryz sat
38	34	7.5	956	8 AK058943	AK058943 Oryz sat
39	34	7.5	1344	6 BD243128	BD243128 Hyperdens
40	34	7.5	1344	6 BD243666	BD243666 Hyperdens
41	34	7.5	1344	6 AR403552	AR403552 Sequence
42	34	7.5	1344	6 AX287002	AX287002 Sequence
43	34	7.5	1344	6 AX343147	AX343147 Sequence
44	34	7.5	1344	6 AX370155	AX370155 Sequence
45	34	7.5	1344	6 AX536272	AX536272 Sequence

ALIGNMENTS

RESULT 1
LOCUS 144909
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmone11
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES
source
Location/Qualifiers
1..456
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 1,7e-120;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCGAGCATTCGACGATCTAGTTCTGGCAGTGGCTGGCT 60
Db 1 ATGAACCTTTTAAAGTGCGAGCATTCGACGATCTAGTTCTGGCAGTGGCTGGCT 60
QY 61 GGGCTGCTCCACATAGGGGCGGCGGCTAATCAATACCGGCGGCAATAGTTCCGCG 120
Db 61 GGGCTGCTCCACATAGGGGCGGCGGCTAATCAATACCGGCGGCAATAGTTCCGCG 120
QY 121 CCGGACTCAACGTTGACATTTTATAGTACGTTCCGCTAACGCTGCTTGTCTGCA 180
Db 121 CCGGACTCAACGTTGACATTTTATAGTACGTTCCGCTAACGCTGCTTGTCTGCA 180
QY 181 AGCGATGCCGCTAATCTGAAACGACCATTCACAGAGCGGTTATGTAACGCGCCGAT 240
Db 181 AGCGATGCCGCTAATCTGAAACGACCATTCACAGAGCGGTTATGTAACGCGCCGAT 240
QY 241 GTAGCGCAGGGTCCGATAATAGTACTATTGAATCTGACATGCTGCAATGCTTCAAAATAT 300
Db 241 GTAGCGCAGGGTCCGATAATAGTACTATTGAATCTGACATGCTGCAATGCTTCAAAATAT 300
QY 301 GCCACCATGACCATGTAACCGCTAATAAATCTCGATATTATGTCGCAATACGCGCGT 360
Db 301 GCCACCATGACCATGTAACCGCTAATAAATCTCGATATTATGTCGCAATACGCGCGT 360
QY 361 AATAACGCCGCGGTGTTATCAGACCGCATGTTATCCAGGCTATGTCGTCAGATT 420
Db 361 AATAACGCCGCGGTGTTATCAGACCGCATGTTATCCAGGCTATGTCGTCAGATT 420
QY 421 GGTTTTGGCAACACGCCGCTAACCGATATTA 456
Db 421 GGTTTTGGCAACACGCCGCTAACCGATATTA 456

RESULT 2
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbria-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION U43280.1 GI:1184712
VERSION 1
KEYWORDS
SOURCE *Salmonella enteritidis*
ORGANISM *Salmonella enteritidis*
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; *Salmonella*.
REFERENCE
AUTHORS Doran, J.L., Collinson, S.K., Burian, J., Santos, G., Todd, E.C., Munro, C.K., Kay, G.W., Baner, P.A., Peterkin, P.I., and Kay, W.W.
TITLE DNA-based diagnostic tests for *Salmonella* species targeting agfA, the structural gene for thin, aggregative fimbriae
JOURNAL J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
MEDLINE 94013373
PUBMED 8104955
REFERENCE
AUTHORS Collinson, S.K., Cloutier, S.C., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE *Salmonella enteritidis* agfBAC operon encoding thin, aggregative fimbriae
JOURNAL J. Bacteriol. 178 (3), 662-667 (1996)
MEDLINE 96146512
PUBMED 8550497
REFERENCE
3. (bases 1 to 2067)

AUTHORS Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg., Victoria, BC V8W 3P6, Canada
FEATURES
source
Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrC"
571..576
598..603
696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/transl_table=11
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKKLLPFWLFTIGAPGATATNYDLARSEYNPAVELSSSEN SLSQAVGTNSARVROGSKLSVISOEGANNRAKDYDGNTPATIEOTGNANDA 696..752
/gene="agfB"
/53..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids; fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=11
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLLKVAAPAAIVSGSALAGVVPOMGGGNNNGGNSGPDST LSIYQGSANAALALQSDARKSETTIGSGYAGADVGCGADNSTIELTNGRRNAT IDQWNAKNSDITGVQYGNNALVNGTASDSSVWVRQVGNATNATNQT"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=11
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MTLLLAALASNOITFTTQGGDIYTVIPQVTLNBPVCQVOIT SVYDVGQSGSHTQOKTSLPANQPIELSRISVNTSSSDSVKIIYTVDSGSLHSQQ WPSAQ"

ORIGIN

Query Match 100.0%; Score 456; DB 1; Length 2067;
Best Local Similarity 100.0%; Pred. No. 1.8e-120;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	ATGAACTTTAAAGTGGCGCATTCGACATCGTAGTTTCTGGCAGTCTTGCT	60	
Db	1193	ATGAACTTTTAAAGTGGCGCATTCGACATCGTAGTTTCTGGCAGTCTTGCT	1252	
Qy	61	GGCGTCGTTCCACATGGGGGGGGGGTAAATCATATACGGGGCGGCAATGATTCGGC	120	
Db	1253	GGCGTCGTTCCACATGGGGGGGGGGTAAATCATATACGGGGCGGCAATGATTCGGC	1312	
Qy	121	CCGAGCTCAACGTTAGCATTTATCAGTACCGTTCCGTTACCGTCCGCTCTCGCA	180	
Db	1313	CCGAGCTCAACGTTAGCATTTATCAGTACCGTTCCGTTACCGTCCGCTCTCGCA	1372	
Qy	181	AGCGATGCGGCTAAATCTGAAAACGACATTACCCAGACGGTTATGTTACGGCGCGAT	240	
Db	1373	AGCGATGCGGCTAAATCTGAAAACGACATTACCCAGACGGTTATGTTACGGCGCGAT	1432	
Qy	241	GTAGCGCAGGGTGGGATATAGTACTATTAAGTCACTGAGATCGTTTCAGAAATAT	300	
Db	1433	GTAGCGCAGGGTGGGATATAGTACTATTAAGTCACTGAGATCGTTTCAGAAATAT	1492	
Qy	301	GCCACATCGACGACGTGGAACGCTAAAACTCCGATATTACTGTCCGCCAATAACGGCGGT	360	
Db	1493	GCCACATCGACGACGTGGAACGCTAAAACTCCGATATTACTGTCCGCCAATAACGGCGGT	1552	
Qy	361	AATACCGCGCGCTGCTGTTATCAGACCGCATCTGATTCACGCTAATGTCGCTCAGGTT	420	
Db	1553	AATACCGCGCGCTGCTGTTATCAGACCGCATCTGATTCACGCTAATGTCGCTCAGGTT	1612	
Qy	421	GGTTTGGCAACAGCGCAGCGGCTAACGATATTAA	456	
Db	1613	GGTTTGGCAACAGCGCAGCGGCTAACGATATTAA	1648	
RESULT 3	STA2301	5103 bp	DNA	linear
LOCUS	STA2301			BCT 15-NOV-2000
DEFINITION	Salmonella typhimurium csgG, csgF, csgE, csgD, csgB, csgA, and csgC genes.			
ACCESSION	AJ002301			
VERSION	AJ002301.1	GI:2739232		
KEYWORDS	csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene; csgG gene.			
SOURCE	Salmonella typhimurium			
ORGANISM	Salmonella typhimurium			
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.			
AUTHORS	Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.			
TITLE	Curl fimbriae are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation			
JOURNAL	J. Bacteriol. 180 (3), 722-731 (1998)			
MEDLINE	98117058			
REFERENCE	9457880			
AUTHORS	2 (bases 1 to 5103)			
TITLE	Romling, U.			
JOURNAL	Direct Submission			
FEATURES	Submitted (29-OCT-1997) Romling U., Department of Bacteriology, Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN			
SOURCE	Location/Qualifiers			
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			
	/protein_id="CAA05311.1"			
CDS				
	1..5103			
	/organism="Salmonella typhimurium"			
	/mol_type="genomic DNA"			
	/strain="SR-11"			
	/db_xref="taxon:602"			
	<1..332			
	/note="ORF179a"			
	/codon_start=2			
	/transl_table=1			
	/product="hypothetical protein"			

```

gene
  3344..3376
  /gene="csGBA"
  /protein_id="CA05319.1"
  /db_xref="GI:2739241"
  /db_xref="SWISS-PROT:054295"
  /translation="MSVKKKIPAIKICAFHISAVGQQTIVGGVHFRGAIIVPEPL
  CDS
  3469..3924
  /gene="csGB"
  /function="nucleator"
  /evidence=experimental
  /codon_start=1
  /translation="MKKKLLPMLTIIAGGIAIATNTYDLARSEYFANVELSKSFN
  sig_peptide
  3469..3531
  /gene="csGB"
  3532..3921
  /gene="csGB"
  /product="CsGB protein"
  /function="nucleator"
  3966..4421
  /gene="csGA"
  3966..4421
  /gene="csGA"
  /function="fiber subunit"
  /codon_start=1
  /translation="MKLKVAFALIVVGSALAGVVPOMGGGNNHNGSSGPPST
  sig_peptide
  3966..4025
  /gene="csGA"
  4026..4418
  /gene="csGA"
  /product="CsGA protein"
  /function="fiber subunit"
  4483..4809
  /gene="csGC"
  4483..4809
  /gene="csGC"
  /codon_start=1
  /translation="MHTLILALASQITFTTQGGDIYTVIPQVTLNPEVCQVQIL
  mat_peptide
  4483..4533
  /gene="csGC"
  4534..4806
  /gene="csGC"
  /product="CsGC"
  4940..5103
  /gene="ORF103"
  4940..5103
  /gene="ORF103"
  /codon_start=1

```

ORIGIN

Query Match 98.2%; Score 448; DB 1; Length 5103;
 Best Local Similarity 98.9%; Pred. No. 3,96-118;
 Matches 451; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACGACATCGTATGTTCTGCGAGTCTCGGT 60
  3966 ATGAACCTTTAAAGTGGAGCATTCGACGACATCGTATGTTCTGCGAGTCTCGGT 4025
QY 61 GCGCTCGTTCACAAATGGGCGGCGGCGGCTAATCATTAACGGCGGCAATAGTCCGCG 120
  4026 GCGCTCGTTCACAAATGGGCGGCGGCGGCTAATCATTAACGGCGGCAATAGTCCGCG 4085
QY 121 CCGACTCAAGTTGAGCATTTATCATGACGCTCCGCTAAGCTGGCGCTGCTCGAA 180
  4086 CCGACTCAAGTTGAGCATTTATCATGACGCTCCGCTAAGCTGGCGCTGCTCGAA 4145
QY 181 AGCGATGCCCGTAAATCTGAACGACATTAACCCAGACGGTTATGTAAAGCGCGCAT 240
  4146 AGCGATGCCCGTAAATCTGAACGACATTAACCCAGACGGTTATGTAAAGCGCGCAT 4205
QY 241 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
  4206 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 4265
QY 301 GCCACCATGACAGTGAAGCTAAATCTCCATTTACTGTGGCCAAATACGGCGGT 360
  4266 GCCACCATGACAGTGAAGCTAAATCTCCATTTACTGTGGCCAAATACGGCGGT 4325
QY 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATCCAGCGAATGGCGCTAGTT 420
  4326 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATCCAGCGAATGGCGCTAGTT 4385
QY 421 GATTGGCAACAAGGCGGCGGCTAATCAATTTAA 456
  4386 GATTGGCAACAAGGCGGCGGCTAATCAATTTAA 4421
Db

```

RESULT 4

AE008749

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITILE

JOURNAL

MEDLINE

PUMED

REFERENCE

AUTHORS

CONSRTM

JOURNAL

Submitted (29-MAR-2001) Genome Sequencing Center, Department of

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI433283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/sofflab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
Reguondb;
http://kitch.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES

SOURCE

location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

434..1308

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; Reguondb:STMS1H001398"

454..1308

/gene="phoH"

/note="similar to E. coli Phob-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="Phob-dependent ATP-binding pho regulon
component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRQVENAGDVYQLKLSNGRQAVIKAREAKRVLRRDS
RSHKRESEVSTLVQMGVGEALIGMARDSDPTSVKARNEQAHYLNALISKOIIPAT
YDVLKRLGASFMQYCLRPEIKVEIAPPAVGRGTENAVVILDEAONVTAQMKRF
LTRGENVTVTVNDITQCDLPRGVRSGSLDARFREDEMVGIYHNRKDCVTSALC
ORTIHAYS"

complement(1414..2303)

/gene="STM1127"

complement(1414..2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcripional regulator"

/protein_id="AAL20056.1"

/db_xref="GI:16419643"

/translation="MEOPRRLKPKGLDITGAMQKSLTRASQRIADYIAPPOVTO
SSIALDSRETQAGATVIRFCRTLGKGFQDFKNDLIELATTSDDSSPLDVEE
SDDHAIGLKQNTISNVLSETNLDMQVLGVDAALRHCHSYIRFGVSGGTALD
MKHLMRIIGRGDAVSNHFMVQATLLKADVAMGVSHSGTSPETVHSIRARQGA

RBS

TTVATHTNIGSPICEADPCLINGNRQMLQGDSTGRKAQLFVFDLLYTLVQSSPE
QARSKRTNMAIDMTX"
complement(2298..2303)

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

CDS

RBS

gene

complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; Reguondb:STMS1H001399"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MTHSRGIVYVLYVFGYLLAMLVGVYFSRROKADDPYGGGR
VPGMAGVSPVATTLISITFMSIPAKAFSDMTDFIIGOVLAIALPLVYFVYPPRK
LKTSAVEYELARDVRCRLPASMPLPHIGRAIITPLVLAIRPAIDPILVIL
LISVCIITVMGIGBGVITVDVIGLLSGSALITFVCLKVGQGIDELFTVQQA
DKPEPAPQFMWSWTESTVPLMIGFLPANIQQFTASDVVQVRYVTDSIEBTKTLLT
NAXLAVAVIPVFFPAIGSALFVYQHQQLPAGNTGIIPLFVTEMVPGIACIIIA
AIFAAQSSISSISINSISCSFNSDIYORLSHKRTPRMKRIAKLVILVAGLISAS
VLIWMADESEIWDAPNSLIGMGPMGLFMIGIFPRANAGSAVIGIITISVITVLA
RYATDLPFFYGVGSLSVYISGIVRAPLAPAPPLTDEKPEKTL"
complement(4088..4088)
/gene="STM1128"
/note="putative RBS for STM1128; Reguondb:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSLLARLEQSVHENGGLIVSCQVPSGPMDEPPIVAAQAAS
AGAVAVRIGENRTVRPHLSVPIIGIRDLGSGVRTTPYQVDAALAAQADII
APGAVPSRSRVDDSLTRIRHGLIAMAACSTVNEGISCHOKIEITIGTSGVTP
ITVPEPDLAVTQLSHAGRCVIAEGRYNTPALANALIEHGAIVTGSALTRIHICQ
WFSHAVER"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; Reguondb:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQMKRPFKMMTIALCSVAITTSGYATLLPDIPBLKNGG
AIDNNGVITGLGTAGTSWYKIDKKQKHMERIKSPGCAREQSVFVFLDELTVYG
GVGKNSKESPLQVYSVYKYSVPAKNTQKQDTISPVGLTGTGKLETNVLITGVN
EHI.FDKYFIDIAAADADESKNKVINYFVKPAKDYEPKIVFVYNAKENVTKNAGEL
PGAGTSSSVMMNNFLMLINGELKPLRVDVYRAMMDNDKLTMLKNSQLPSPGSG
QOQMGKAGFSGYSHGVLLVGGANFPKAXNTNGKRYSHGINKKRRDRVYGLINGH
WOMGKMKPGLGVGSVSYSDVFLVIGENAKKXPVSVTFMRDNLILIK"
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; Reguondb:STMS1H001402"

```

6805. .7497
/genre="STM1131"
/notes="similar to E. coli orf, hypothetical protein
(AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
identity in aa 4-239"
/codon_start=1
/transl_table=1
/product="putative outer membrane protein"
/protein_id="AAL20062.1"
/db_xref="GI:1641964"
/translation="MKIKLIVGNVSFIASSYLQATLVDREHYADRTIRNKRIAL
IKELNNGIGFVDAVSXSGVDGSDGKHLSDLVANAIEGVSYKVTNDVLDPGII
RESGSDITVYPRYLRGGQVPEFSGVYAGRRRYVAKTANVSDDSKTRPPTIYGIYF
DELKLENNFTMMDSDDIQRPNKKITNEYHNVALAMKLKNSFTIYEVGVNAVRNNTDR
QTRYVVGQIYHF"
7766. .9060
/genre="STM1132"
7768. .7773
/genre="STM1132"
/notes="putative RBS for STM1132; RegulonDB:STMS1H001403"
7780. .9060
/genre="STM1132"
/notes="similar to E. coli putative transport protein
(AAC77235.1). Blastp hit to AAC77235.1 (425 aa), 57%
identity in aa 25-418"
/codon_start=1
/transl_table=1
/product="putative sugar transport protein"

```

Query Match	98.2%	Score 448;	DB 1;	Length 22411;
Best Local Similarity	98.9%;	Pred. No. 4.3e-118;		
Matches 451; Conservative	0;	Mismatches		

Qy	1	ATGAACTTTTAAAGTGGAGCATTGCGACGATCGTAGTCTTGCGAGTGTGGCT	60
Db	17769	ATGAACTTTTAAAGTGGAGCATTGCGACGATGTAGTCTTGCGAGTGTGGCT	1782
Qy	61	GGGCTCTTCCACATATGGGGCGGGCGGTATCATTAACGGCGCGGCATATGTTCCGGC	120
Db	17829	GGCGTCTTCCACATATGGGGCGGGCGGTATCATTAACGGCGCGGCATATGTTCCGGC	1788
Qy	121	CCGAGCTCAACGTTGAGCATTTATCAGTACGGTTCGGTCAACGTCGGCTGTCTCGAA	180
Db	17889	CCGAGTTCACGTTGAGCATTTATCAGTACGGTTCGGTCAACGTCGGCTGTCTCGAA	1794
Qy	181	AGCGATGCCCGTAAATCTGAACGACCATTAACCCAGACGGTTATGTACGGCGCGAT	240
Db	17949	AGCGATGCCCGTAAATCTGAACGACCATTAACCCAGACGGTTATGTACGGCGCGAT	18001
Qy	241	GTAGGCGAGGGTGGCGATATATGTACTATTGACGACTCAGATAGTTCACAGAAATAT	300
Db	18009	GTAGGCGAGGGTGGCGATATATGTACTATTGACGACTCAGATAGTTCACAGAAATAT	18066
Qy	301	GCCACCATTCGACCACTGGAGCGCTAAAAATCCCGATATTACTGTGGCCAAATACGGCGGT	360
Db	18069	GCCACCATTCGACCACTGGAGCGCTAAAAATCCCGATATTACTGTGGCTCAATACGGCGGT	18128
Qy	361	AATTAACCGCGGCTGTGTTAATCAGACCGCATCTGATTCACGCGTAATGGGCGCTCAGGTT	420
Db	18129	AATTAACCGCGGCTGTGTTAATCAGACCGCATCTGATTCACGCGTAATGGGCGCTCAGGTT	18188
Qy	421	GGTTTGGCAACAACGCCAGCGGCTTAACCAAGTATTTAA	466
Db	18189	GGTTTGGCAACAACGCCAGCGGCTTAACCAAGTATTTAA	18224

RESULT 5	LOCUS	DEFINITION	DNA	BCR
AL627269	AL627269	Salmonella enterica serovar Typhimurium complete chromosome, segment 5/20.	linear	04-JUN-2003
ACCESSION	AL627269	ALU51382		strain CT18,
VERSION	AL627269.1	GI:16502231		

KEYWORDS

ORGANISM
Salmonella enterica subsp. enterica serovar Typhi
Salmonella enterica subsp. enterica serovar Typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales
Enterobacteriaceae; Salmonella
1 (bases 1 to 254050)
Parthill, J., Dougan, G., James, K. D., Thomecz, N. B., Salazar, A. L.,
REFERENCE
AUTHORS

AUTHORS

Walt, J., Churcher, C., Mungall, K. L., Bentley, S. D., Pickard, D., Sebahia, M., Baker, S., Basham, D., Brooks, K., Chillingworth, T., Comerford, P., Cronin, A., Davis, P., Davies, R. M., Dow, L., Farrar, J., Felwell, T., Hamlin, N., Haque, A., Hien, T. T., Hiley, N., Jags, K., Krogh, A., Larsen, T. S., Leather, S., Mole, S., O'Gaora, P., Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S. and Barrett, B. G. Complete genome sequence of *Salmonella enterica* serovar Enteritidis phage type 4 isolated from a human in 2003, and comparison with other *Salmonella* genomes

TITLE

JOURNAL
OF
CLINICAL
MICROBIOLOGY
Volume 41, 1999, 1000-1004
DOI: 10.1093/jcm/41.5.1000
Printed in the USA
© 1999 American Society for Microbiology
0893-9726/99/051000-05\$05.00/0

MEDLINE
DITW/MT

REFERENCE

REFERENCE

AUTHORS
TITLE
JOURNAL

COMMENT

Details of *S. typhi* sequencing at the Sanger Centre are available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

FEATURES

Source

```

/organism="Salmonella enterica subsp. enterica serovar
Typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
181..2793
gene

```

CDS

```

/gene="STY1078"
/note="Orthologue of E. coli pepN (AMPN_ECOLI); Fasta hit
to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1

```

misc

misc_

/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:O82759"
/db_xref="SWISS-PROT:O82759"
/translation="MYEFVKALFQDLPERAHETFEQLRRTITGTPLEALVRQVP
KPTCMGLTFKGLAGLDKDGSCIALGAMGSGLEIGTTPRPOGKDLPR
LVDAEGLINRMGFNNIGVDNLVENYKKAHPGIIIGINIKNDTPVENGKDYLCME
KVYAYGIIAIIINISPNTRGLTQYGLALDLITLTKRNDLQVTHKRYPAVKI
APDCBEELIVADSLRNIDGVATATATDRLSDRSVQMKRCQGTGSLSPDKST
EIRRLSELKGLPIIGVGIDSVIAREKIAAGATLVQISGTFIKGPLIKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBM_ECOLI; Faeta hit to
YCBM_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:O82758"
/translation="MMRIKPDNDMRYDEEDHDMMLDIANGMLFRSFRSKMLTPDA
FCPTGFCVDDAALYFSEBEKCRDELTKQRAELVINALVAIRYLKQPMPSMHPVA
GEMTPTGTGAASWLSDTAQQVNLVVEBENALCLLAQPGVVIAGRTVQLDALIK
IMNDRLKRVQVCHSFSLQAV"
complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:O82757"
/db_xref="SPTREMBL:O82757"
/translation="MATLSRLFIHPVKSMRGITGLTHALADISGLAPRIEMVTESDGT
FITARQPPQWRFTSPPLHDLHLTAPDGSSALVRFDPFQDAPTEWGMFAPARA
PTAINOMLSGFSEFSDVOLCWYGPOLTRVRKHNAVLPFDVYKPSRCITPTVSEKQ
KHPSGEPLATIQARTIADNGDDVDFQQLIARNSGVIRVGEVEILATAPKAYGATT
LDBSVTEPKHPSGVITIDWQGTFCGNQOVLLEOLFENQGIIRIPYSCRAGICGCRIR
LLEGEVSPKHSAMGDDGTTILSCVVKTXLRLEN"
complement(4737..4955)
/gene="STY1081"

/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY_HAEIN (p43945) Faeta
scores: E() 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:O82756"
/db_xref="SPTREMBL:O82756"
/translation="MNSLFASTAGLSEBLKIELEKAGVCCQVYVGGVHFGQDTRLI
YQSLMWSRLASRIILPMGECKVYSDDLDFGVQAIMTEIFNPATFVHPSGLNDTI
RNSQYGMKXVDAIVDAFTTRKNLPRPVNDRESPLRIINVLNKTASTIALDLSGDLH
LRGYRDRTGAPLKEITLAAIIVMSGQPGTLPDPMCGSGTLLIEMAMVTDRAFG
HRGMGFSGMARHDEALWQSVKAEQOTRARKGLAESHFYGSDADAVIRASNAR
RAGIGLITFEVQVADQSLNPLPKPGYGVISNPVGBRDLSDPAILALHSLGTYM
NORGAMLSLFSASPDLSQILHADQFPAKNGPLDVCQKNTIATETADSKPTVA
EDVANNRIRKMLKLEKARQEGIECVLYADLDEYVAVADRVDMVATQSBAPKTV
RLMVLITDYLDTGLFLDRIARMLGMSKIDKIDFYSASVAGSAGANNTT
VDMRTYLEMAERNLRUNGSGRAHRLIQQDCGLMLEANQOPLFIPIDEPTFSNKR
MESFPDVRQHVLMKRLKRLKRGITIMSNKRGFRMDLEGLAEGLTQBITQKT
LSPDFARNRQIHNCMLRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIR_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa
overlap

Query Match 97.9%; Score 446.4; DB 1; Length 254050;
Best Local Similarity 98.7%; Pred. No. 1.5e-117;
Matches 450; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy	1	ATGAACTTTTAAAGTGCAGCATTCGGCAGCAATCGAGTTTTCGAGTGCCTGGCT	60
Db	88906	ATGAACTTTTAAAGTGCAGCATTCGGCAGCAATCGAGTTTTCGAGTGCCTGGCT	88965
Qy	61	GGCGTCTTCCACAAATGGGGCGGGCGGTATCATAGCGGGCGCAATAGTCCGGC	120
Db	88966	GGCGTCTTCCACAAATGGGGCGGGCGGTATCATAGCGGGCGCAATAGTCCGGC	89025
Qy	121	CCGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTTCTCGAA	180
Db	89026	CCGACTTCCACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTTCTCGAA	89085
Qy	181	AGCGATCCCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTGAACGGGGCGAT	240
Db	89086	AGCGATCCCGTAAATCTGAAACGACATTAACCAAGCGGTTATGTGAACGGGGCGAT	89145
Qy	241	GTAAGCCAGGGTGGGATTAATAGTACTATTGAATGACTGCAATGGTTTCAGAAATAT	300
Db	89146	GTAAGCCAGGGTGGGATTAACAGTACTATTGAATGACTGCAATGGTTTCAGAAATAT	89205

```
Salmonella typhi CT18"
/codon_start=1
/translate table=11
/product="flagellar hook-associated protein 3"
/protein_id="AA059360.1"
/db_xref="GI:29137799"
/translation="KRLSTOMMYEQNNSGILTNSQAEWMKLGQMSYTGKRVTPSDPP
AASQAVLTLDQAGNSQALARFATQKVSLEESVLSQVTAIQTQKQETVYAGSTT
SDDRSALVLTLDQAGIRDLQNLNINASTGNSGRVIFAGKVEAAFPQATQGYHGKSKT
TQGVDEARVMTVGHGTGAQIFNSITSNVAPDPSDSDQKFWLTLTALIKLTPYBGA
DKVEKAAAIADIKTRGKINSINLVLTFRVRLTETQI.SRT.SLT.SCT.CGCTG
```

```

CDS
    /locus_tag="C1737"
    complement(1812..3473)
    /gene="flgK"
    /locus_tag="C1737"
    /note="corresponds to STY1222 from Accession AU513382
    Salmonella typhi CT18"
    /codon_start=1

```

gene	****CCTAATGAGTGGTGTGTTTAAACCAATGTAATTCATTATGTTT Complement(3538, .4488)
CDS	/gene="figU" /locus_tag="t1738" complement(3538, .4488) /gene="figU"

Salmonella typhi CT18"

/codon_start=1

/transl_table=11

/product="flagellar protein FlgG"

/protein_id="AA069362.1"

/db_xref="GI:29137801"

/translation="MTGDKRLTASAMPAQSIINELKAKAGODPAANIRPVAROVEGK
VOMKIKSRMALPKDGLPSSDQTRLVYSYMDQALAOQMTAKGIGLADMMVKQGTGG
TMEPLDDAPVPLKFLSETNYSYONALQTLVQRKALPKTPSSDPLSGDSKDLPLAT
LPLALASQSGVPHHLLAQAALSGMGCRDQILBNBGRPTNFVGVAATSMGKPLAT
LTITVEENGSGAKVKAPRVVSVYLEALSDVLTALLTRPAAVTTAATAEQCAVAL
NAGADDPNVPARKLASMTQQLKAMSEKSVKTSYANLDNL"

complement(4488..5591)

/gene="flgI"

/locus_tag="t1739"

complement(4488..5591)

/gene="flgI"

/locus_tag="t1739"

/note="corresponds to STY1220 from Accession AL513382:

Salmonella typhi CT18"

/codon_start=1

/transl_table=11

/product="flagellar P-ring protein precursor"

/protein_id="AA069363.1"

/db_xref="GI:29137802"

/translation="MVVFKLAGIVLAVVATLAHERIDLTSGVQRENSLIGYGLVGLD
VLDSTGQOTQTPFTQTLNNMLSQSLGIVPVTGNMQLKNVAVMVATSPFAPRQD
QITDIVVSSMAKASLKGILLMTPLKGVLAQGNILIVGAGASAGGSSVQVQV
QNGIRVTNGAILIRELPFGAGNTINQLDNDEFTMAOQITDLAINARQVGSATL
DKRTVQVRPGNSSVYRFLADLIQNMVEVAVTPQDAKAVYNRSRTSVVNRPRVTPDCT

gene
BALNAGATPMDLMSILQSMOSAGCLRAKLEII"
complement(5597. .6295)
/gene="flgH"
/locus_tag="c1740"
complement(5597. .6295)
/gene="flgH"
/locus_tag="c1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQKTAALHAYPWPALMTVLTGCCAMIPAKPLVQGAATTQAPLPGPV
PVANGSIFQSAQPINYGYPLFEDRPNIDTLTIVQENVASAKSSAANARDKGT
SFGPTVRYVLYQGLFGNSRADMEASGSGFNGKGANASNTFSGTLTIVQVLANGN
LHVGEKQIAINOGTEFRFSGVVPRTISGNSVPSYQVADARIEYGVNGYINQON
MGWLORFPLNLSPM"
complement(6350. .7132)
/gene="flgG"
/locus_tag="c1741"
complement(6350. .7132)
/gene="flgG"
/locus_tag="c1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISLWTAKTGLDAGQNNMVIANNLANVSTNGEKROPAFEDL
LYQIRQGAQSSQETLPSGLQIGTVRPATERLHSGNLSTQNNKDAIKQGGF
FQWMLPDGSAVTRDGSFQVNDQGLTVAGSFQVQPAITIPANALSTIGRDGVSVT
QOQGAAPVQVQLNLTFTFMDTGLSEIGENIYETQSSGAPSESTPGANGAGLYQGY
VETSNNVAEEVNMIVQRAYEINSKAVSTTDQWLQRLTQL"
complement(7146. .7901)
/gene="flgF"
/locus_tag="c1742"
complement(7146. .7901)
/gene="flgF"
/locus_tag="c1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/transl_table=1
/codon_start=1
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHALITRMGAASQTLANQAVTASNLANASTPGFRAQLNALRAV
PVDGLSLATRLTLVASTPGADMTQGLDYTSRPLDVALAQDQGLVQADAGEGTRN
GNIQVPTGQGLTIQHPVIGSGPITVEGSEITLADGITSALNPGSPPTVAEVR
LKLVAEGNEVORSDDLFLRLAEQAERGVLADPSIRIMSGVLGSGSNVKEPYEAMT
DMINARREFEMQKVTISVDNEGRANLLMS"
complement(7922. .9133)
/gene="flgE"
/locus_tag="c1743"
complement(7922. .9133)
/gene="flgE"
/locus_tag="c1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/transl_table=1
/codon_start=1
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAVSGLMAATNIDVIGNNANGATGFKSGTASPADMPA
GSKVGLGVKAVGITDDFTDGTITNGRLDVAISQNGFRLDSDSGSVFYSRNGQFKL

Query Match 97.9%; Score 446.4; DB 1; Length 301983;
Best Local Similarity 98.7%; Pred. NO. 1.5e-117;

Matches 450; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGACATTCGACCAATCGTAGTTTCTGGAGTCTTGCT 60
DB 37310 ATGAACTTTTAAAGTGGACATTCGACCAATCGTAGTTTCTGGAGTCTTGCT 37251
QY 61 GGGCGCTTCCACAAATGGGGGGGGGGGGTAAATCAATACGGCGGGGGAATGTCGGC 120
DB 37250 GGGCGCTTCCACAAATGGGGGGGGGGGGTAAATCAATACGGCGGGGGAATGTCGGC 37191
QY 121 CCGACTCAAGCTTGAAGATTATTCAGTACGGTTCGGTAAACGTCGGCTTCTGCA 180
DB 37190 CCGACTTCAAGCTTGAAGATTATTCAGTACGGTTCGGTAAACGTCGGCTTCTGCA 37131
QY 181 AGCGATCCCGTAAATCTGAAACGACCATTAACCAAGCGGTTATGTTAACGGCCCGAT 240
DB 37130 AGCGATCCCGTAAATCTGAAACGACCATTAACCAAGCGGTTATGTTAACGGCCCGAT 37071
QY 241 GTAAGCCAGGGTGGGATTAATGACTATGTAAGTGAAGTCAAGTTCAGAAATAT 300
DB 37070 GTAAGCCAGGGTGGGATTAATGACTATGTAAGTGAAGTTCAGAAATAT 37011
QY 301 GCCACATCGACCAAGTGAACGCTAAAACTCCGATATTACTGTCGGCCAAATACGGCGGT 360
DB 37010 GCCACATCGACCAAGTGAACGCTAAAACTCCGATATTACTGTCGGCCAAATACGGCGGT 36951
QY 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCACGCTAATGTCGTCAGGTT 420
DB 36950 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCACGCTAATGTCGTCAGGTT 36891
QY 421 GATTTGGCAACACGCGCAGCGCTTACAGTATTA 456
DB 36890 GATTTGGCAACACGCGCAGCGCTTACAGTATTA 36855
RESULT 7
STAGFBA 1048 bp DNA linear BCT 26-JAN-1998
LOCUS Salmonella typhimurium agfB and agfA genes.
DEFINITION AJ000514
ACCESSION AJ000514.1 GI:2275119
VERSION agfA gene; agfB gene.
KEYWORDS Salmonella typhimurium
SOURCE Salmonella typhimurium
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE 1
Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
Infect. Immun. 65 (12), 5320-5325 (1997)
JOURNAL
MEDLINE 98053981
PUBMED 9393832
REFERENCE 2 (bases 1 to 1048)
Sukupolvi,S.S.
Direct Submission
Submitted (14-JUL-1997) Sukupolvi,S.S., Medical Biochemistry,
University of Turku, Kiinamyllykatu, 20520, FINLAND
FEATURES
source
1. 1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14. .553
/gene="agfB"
14. .553
/gene="agfB"
/codon_start=1
/transl_table=1
/product="agfB protein"
/protein_id="CAA04150.1"

/db_xref="GI:2275120"
 /db_xref="SPTREMBL:O33801"
 /translation="MKRKLFLMMLTLIGSPIATATNYDLARESEVFAVNEISKSPNO
 AATIGVDNDSARVHESKILSVISGERGNNAKIDQANVPAITETGTANADA
 SISOASVGNSSAISAIIQKSGNKANITQYGTQKTAIVVVKOSMAIQANITQYGTQ
 KTAIVVVKOSMAIRVYOR"
 593..1048
 /gene="agfa"
 593..>1048
 /gene="agfa"
 /codon_start=1
 /transl_table=11
 /product="Agfa protease"
 /protein_id="CAA04151.1"
 /db_xref="GI:2275121"
 /db_xref="SPTREMBL:O33802"
 /translation="MKILKVAFAIVVSGAVVPPQWGGGNGHNGSSGPDST
 LSIYQGSNNAALAIQDARKSETTITQSGYNGADVGCGADNSTELIENGPRNNAT
 IDQWAKNSDITVGYGSGNNAALVNTASDSVVRVYGFGNNAPANQYN"

gene
 CDS

ORIGIN

Query Match 96.6%; Score 440.6; DB 1; Length 1048;
 Best Local Similarity 98.0%; Pred. No. 4.9e-116;
 Matches 446; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGAGCAATCGTACTTTCTGGAGTGTCTGGCT 60
 DB 593 ATGAACCTTTAAAGTGGAGCATTTGCGAGCAATCGTACTTTCTGGAGTGTCTGGCT 60
 QY 61 GGCCTGCTTCCAAATGGGGCGGGCGGCTAATCAATACGGGGCGGCAATGTTCCGGC 120
 DB 653 GGCCTGCTTCCAAATGGGGCGGGCGGCTAATCAATACGGGGCGGCAATGTTCCGGC 120
 QY 121 CCGGACTCAAGTGGAGCATTTATCAATAGCTGCTCGCTAACGCTGCTGCTGCA 180
 DB 713 CCGGACTTCCAGTTGAGCATTTATCAATAGCTGCTCGCTAACGCTGCTGCTGCA 180
 QY 181 AGCGATGCCGTAATCTGAAAGCAATTAACCGAGGCGTTATGTAACGGCGCGAT 240
 DB 773 AGCGATGCCGTAATCTGAAAGCAATTAACCGAGGCGTTATGTAACGGCGCGAT 240
 QY 241 GTAGGCGAGGTCGGATTAATGTAATTAAGTACTGACTCAGAAATGTTTCAAAATAT 300
 DB 833 GTAGGCGAGGTCGGATTAATGTAATTAAGTACTGACTCAGAAATGTTTCAAAATAT 300
 QY 301 GCCACCATGACAGTGAACGCTAAAACTCCGATTAATGTCGCGCAATACGGCGGT 360
 DB 893 GCCACCATGACAGTGAACGCTAAAACTCCGATTAATGTCGCGCAATACGGCGGT 360
 QY 361 AATAAGCGCGGCTGTAATCAAGCGCATCTGATTCAGCGTAATGTGCGTCAAGTT 420
 DB 953 AATAAGCGCGGCTGTAATCAAGCGCATCTGATTCAGCGTAATGTGCGTCAAGTT 420
 QY 421 GGTTCGCAACAAAGCGCGGCTAAACGATTA 455
 DB 1013 GGTTCGCAACAAAGCGCGGCTAAACGATTA 1047

RESULT 8 361 bp DNA linear PAT 07-OCT-1997
 LOCUS 144908
 DEFINITION Sequence 56 from patent US 5635617.
 ACCESSION 144908
 VERSION 144908.1 GI:2469621
 KEYWORDS
 SOURCE
 ORGANISM
 Unknown.
 Unclassified.
 1 (bases 1 to 361)
 AUTHOR Doran,J.L., Kay,W.W., Collinson,S. Karen. and Clouthier,S.C.
 TITLE Methods and compositions comprising the agfa gene for detection of
 JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

FEATURES
 source Location/Qualifiers
 1..361
 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN

Query Match 72.8%; Score 331.8; DB 6; Length 361;
 Best Local Similarity 99.4%; Pred. No. 1.2e-84;
 Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 64 GTGCTTCACAAATGGGGCGGGCGGCTAATCATATACGGCGGCAATAGTCCGGCCGC 123
 DB 1 GTGCTTCACAAATGGGGCGGGCGGCTAATCATATACGGCGGCAATAGTCCGGCCGC 60
 QY 124 GACTCAACGTTGAGCATTTATCAATAGTTCGGCTTACGATGCGCTTCTGCAAAAC 183
 DB 61 GACTCAACGTTGAGCATTTATCAATAGTTCGGCTTACGATGCGCTTCTGCAAAAC 120
 QY 184 GATGCCGCTAATCTGAAAGCAATTAACCGAGCGGTTATGTAACGGCGCGATGTA 243
 DB 121 GATGCCGCTAATCTGAAAGCAATTAACCGAGCGGTTATGTAACGGCGCGATGTA 180
 QY 244 GGCAGGGTCGGATTAATGTAATTAATGTAATGTAATGTAATGTAATGTAATGTA 303
 DB 181 GGCAGGGTCGGATTAATGTAATTAATGTAATGTAATGTAATGTAATGTAATGTA 240
 QY 304 ACCATGACACAGTGAACGCTAAACCTCCGATTAATGTAATGTAATGTAATGTAAT 363
 DB 241 ACCATGACACAGTGAACGCTAAACCTCCGATTAATGTAATGTAATGTAATGTAAT 300
 QY 364 AACGCCGCTGTTAATCAGACCGCATCTGATTC 398
 DB 301 AACGCCGCTGTTAATCAGACCGCATCTGATTC 335

RESULT 9 2889 bp DNA linear BCT 24-JUN-2003
 LOCUS CSP515700
 DEFINITION Citrobacter sp. Fec2 csb gene, csb gene and csb gene.
 ACCESSION AJ515700
 VERSION AJ515700.1 GI:11790491
 KEYWORDS csb gene; csb gene; csb gene; curli monomers; regulatory protein.
 SOURCE
 ORGANISM
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 1
 Zogaj,X., Bokranz,W., Nimitz,M. and Romling,U.
 Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 Infect. Immun. 72 (7), 4151-4158 (2003)
 JOURNAL
 REFERENCE
 2 (bases 1 to 2889)
 ROMLING, U.
 DIRECT SUBMISSION
 Submitted (11-NOV-2002) Romling U., Microbiology and Tumorigenology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 LOCATION/Qualifiers
 1..2889
 /organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csb"
 /complement(212..862)
 /gene="csb"
 /function="necessary for transcription of the csbBA
 operon"
 /codon_start=1

gene
 CDS

```

        /trnsl table=11
        /product="putative regulatory protein"
        /protein_id="CAD56670.1"
        /db_xref="GI:31790492"
        /translation="MFNEFHSIHGHTLLLTTPKSLQATALLQHLKQSIALSGKLNIC
        RSLDISSICVLVDMADKGLIHYNODNISRKNNMLKTLILTPDYPADRIENMF
        HINGFYVAADDERVYVNLQGLRGEYFSQKLASLYITHSGRVYNSESSALLTHRE
        KEILNKLIAGSNIENIARSLFISEVTVTHLYNLFKIAVNRQAVSWANDINLR"
        1623..2078
        /gene="csqb"
        1623..2078
        /gene="csqb"
        1623..2078
        /gene="csqb"
        /function="nucleator"
        /codon_start=1
        /trnsl table=11
        /product="nucleation component of curlin monomers"
        /protein_id="CAD56671.1"
        /db_xref="GI:31790493"
        /translation="MKNKLFLMMLTVLGAPGIASATSYDLAHSEYNFAVNELSKSPF
        QAALIGQVGTNNASAKROEGSKLSIVSQEGSSNAKVDQSGAVYFAIAGSGNSNDPA
        SISOYNTNTAIIQKSGSKNRIITQYGTQTAVVVQSGQMAIRVIGR"
        2119..2571
        /gene="csqa"
        2119..2571
        /gene="csqa"
        2119..2571

```

ORIGIN

Query Match	68.1%;	Score 310.4;	DB 1;	Length 2889;
Best Local Similarity	81.6%;	Pred. No. 2.1e-78;		
Matches 372;	Conservative	0;	Mismatches 81;	Indels 3;
				Gaps 1;

QY	ATGAAACCTTTAAAGAGGACAGCAATTCGACCAATTCGATGTTCTGGAGAGCTCTGGCT	60
Db	ATGAAACCTTTTACAGAGGACAGCAATTCGACCAATTCGATGTTCTGGAGAGCTCTGGCT	2119
QY	GAGCGTCGTTCCACATGGGCGCGCGCGGTAATCATTAACGCGCGCGGCATATGTTCCGGC	61
Db	GATTCTGTTCGCAATGGGCGGGGGGGG---TGGCGGCGGCGGGAGACAGCTCCGGC	2179
QY	CCGGACTCAAGSTTGAGCATTTTATCATGTAAGGTTCCGTTAACGCTGGCCTTGCTTCGCA	121
Db	CCGGAAATCGACCCTGAGGATTTTATCATGTCAGCAAGAGTCATTAACGCGCGCCTTGCGTCGAA	2236
QY	AGCGATCCCGGTAATCTGAACAGCACCTTATCCGAGCGGGTTATGGTAAACGCGCGCAT	181
Db	AGCGACCTCTGTAATCTGATACGACACCTTATCATGAAATGGCTTTGGTAAACGCGCGCAGAC	2296
QY	GTAGCGCAAGGTCGGATATAGTACTATTTGAATGACTCATGAAATGTTTCAGAAATAT	241
Db	GTGGGCCAAGGCTCAGATTAACAGCACCATTCATCTGACTCAAAACGCGCTTCAAAAACAAC	2356
QY	GCCACCAATTCGACAGTGGGAAGCGTTAAAACTCCGATATTACTGTGCGGCATTAAGGCGCGT	301
Db	GCCACCAATTCGATCAGTGGGAACGGCAAAATTTGGGCATTTACTGTGAGCCAGATATGGTGA	2416
QY	AATAAGCCCGGCTGGTTAATCAGACCGCATCTGATTTCCAGCGTAAATGGTGGTCAAGTT	361
Db	CATAACCCCGCACTGGTGAACCAACATCGCTCCGATTTCCAGCGTATCTGTGCATCAGGTT	2476
QY	GGTTTGGCAACAACGCAACGCGCTAACCAAGATTTAA	421
Db	GGTTTGGCAACAACGCAACGCGCTAACCAAGATTTAA	2536
		2571

FEATURES	source
LOCUS	2920 bp DNA linear BCT 24-JUN-2003
DEFINITION	Citrobacter freundii csbB gene, csbA gene and csbD gene.
ACCESSION	AF515701
VERSION	AF515701.1 GI:31790495
KEYWORDS	csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation component of curlin monomers; regulatory protein.
SOURCE	Citrobacter freundii
ORGANISM	Citrobacter freundii
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.
AUTHORS	1. Zocag, X., Bokranz, W., Nimtz, M. and Romling U.
TITLE	Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)
JOURNAL	2 (bases 1 to 2920)
REFFERENCE	Romling, U.
AUTHORS	Direct Submission
TITLE	Submitted (12-NOV-2002) Romling U., Microbiology and Tumorsbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
JOURNAL	Location/Qualifiers
FEATURES	1..2920

ORIGIN

and products: all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

Location/Qualifiers

1..10346

/organism="Escherichia coli K12"

/mol_type="genomic DNA"

/strain="K12"

/sub_strain="MG1655"

/db_xref="taxon:83333"

complement(<1..48)

/note="IS3"

<1..5

/note="factor Sigma70; predicted +1 start at 1094689"

5..32

/note="factor Sigma70; predicted +1 start at 1094716"

69..392

/gene="b1028"

69..392

/gene="b1028"

/function="orf; Unknown"

/note="O107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein PAS2_CANAL SW: P43098"

/codon_start=1

/transl_table=1

/product="orf; hypothetical protein"

/protein_id="AAC74113.1"

/db_xref="GI:1787266"

/translation="MEDGVQAMRDYLAGLIDAPSPHOVIANVYAKSEVAPSIKENIS LHLHTVKTSTESLDTPVAFLEISNKKYLGNMLNDPAGVDQGRVHRCRAFSB AKVFK"

275..299

/gene="b1028"

/note="factor Sigma32; predicted +1 start at 1094983"

389..1375

/gene="ycdW"

/note="synonym: b1029"

389..1375

/gene="ycdW"

/function="orf; Unknown"

/note="O328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCGE SW: P47562"

/codon_start=1

/transl_table=1

/product="orf; hypothetical protein"

/protein_id="AAC74114.1"

/db_xref="GI:1787267"

/translation="MIPDYLTFRPOKRNLYITVAIGLLGFYKNAAGFTFPEDEI LIMSAMSGLYOLPSLYALFSLISLGPVFLFRMTSYKQVAVSAYKVKYK FLRSKRPAFLGRFLQEDIDFPSSNALSTFAKPLRLFLILIVIMMTLVSUL ATLVEMRIWEAFLLCYVCLIFYFCRPLMHNDFMACDMYRMRGHFK"

1253..1280

/gene="ycdW"

/note="central position to predicted promoter: -104"

/bound_moiety="Deor predicted site"

1253..1280

/gene="ycdW"

/note="No predicted promoter"

/bound_moiety="Deor predicted site"

1337..1365

/gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1096049"

1494..1745

/gene="b1030"

1494..1745

/gene="b1030"

/function="orf; Unknown"

/note="O83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_PABIT SW: P49065"

/codon_start=1

/transl_table=1

/product="orf; hypothetical protein"

/protein_id="AAC74115.1"

/db_xref="GI:1787268"

/translation="MGGRVCVYTKKKVLSYELFFKXGGGIDSLRSPGCPAHC VHGLSNLSPVDPKSGLLIPVCAIHEKRTFVALL"

1513..1926

/gene="ycdW"

/note="synonym: b1031"

1513..1926

/gene="ycdW"

/function="putative structure; Not classified"

/note="O137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein RS37_YEAST SW: P05759"

/codon_start=1

/transl_table=1

/product="putative ribosomal protein"

/protein_id="AAC74116.1"

/db_xref="GI:1787269"

/translation="WCNIRKKSPTFRSSSINNAVRGGLTFARPAGSPPLTAFVCP TSNMSPVPEPSRGLIPVCAIHEKRTFVALL"

complement(2111..2198)

/gene="serX"

/note="synonym: b1032"

complement(2111..2198)

/gene="serX"

/product="rRNA-Ser"

/function="rRNA; rRNA"

/note="anticodon: GGA; CG Site No. 17878"

2269..2296

/function="rRNA; rRNA"

2269..2296

/note="factor Sigma70; predicted +1 start at 1096980"

2338..2365

/note="factor Sigma70; predicted +1 start at 1097049"

2355..2383

/note="factor Sigma70; predicted +1 start at 1097067"

2393..3370

/gene="ycdW"

/note="synonym: b1033"

2393..3370

/gene="ycdW"

/function="putative enzyme; Not classified"

/note="O325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_COROL SW: P45637"

/codon_start=1

/transl_table=1

/product="putative dehydrogenase"

/protein_id="AAC74117.1"

/db_xref="GI:1787270"

/translation="WRINPAPRPSGVSMDIIFHPTDTOMWTEALKAIPQAVRAM LEDTGMQORSTAVSOVLHMFRRFDDYRLOQSSHSQPLPEYHRDEFTGIGAGVL GSKVQSLQTFPLRCMSRTKSGMPGQSPAGEBSATLSGCRVLTINLPTPTPV GIINQOLEKLPDAGVLTINLARGVAVEDDLAIDSGKVGAMLVFNREPLPESP LMQHPRTITPVAITPBAVAVEYISRTIAQLEKGRVCGQVDRARGT"

3278..3308

/gene="ycdW"

/note="factor Sigma70; predicted +1 start at 1097992"

3383..3410

/note="factor Sigma70; predicted +1 start at 1098094"

3396..3423

Query Match 56.1%; Score 256; DB 1; Length 10346; Best Local Similarity 72.6%; Pred. No. 1,2e-62; Matches 331; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

Oy 1 ATGAACCTTTAAAGTGGACAGATTCGACGAAATCGTAGTTTCGCGAGTCTTGCT 60
 Db 8993 ATGAACTTTTAAAGTGGACAGATTCGACGAAATCGTAGTTTCGCGAGTCTTGCA 9052
 Oy 61 GCGCTCGTTCCACATGCGGCGCGCGGTATCATTAACGCGCGCGAGATGTTCCGGC 120
 Db 9053 GGTGTGTTCTCTAGTACGGCGCGCGGTAAACGCGGTGGTGGTAAATATAGCGGC 9112
 Oy 121 CCGGATCTACGCTTGAGCATTTATCATGATCGGTTCCGTTACGCTGCGCTTCTGCA 180
 Db 9113 CCAATCTGAGCTGAAATTTACAGATGACGATGCGGTGTAATCTGCACTGCTCTGCA 9172
 Oy 181 AGGACGCGCGTAAATCTGAAACGACATTAACGCGGTATGATGATGCGCGCAT 240
 Db 9173 ACTATGCGCGTAACTGATCTGATCTGATCTGATCTGATCTGATCTGATCTGAT 9232
 Oy 241 GTAGCGCGCGTGGGATATATGATCTGATCTGATCTGATCTGATCTGATCTGAT 300
 Db 9233 GTTGTGAGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9292
 Oy 301 GCCACATCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 Db 9293 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9352
 Oy 361 AATAACGCGCGGTGATATGATGATGATGATGATGATGATGATGATGATGAT 420
 Db 9353 GGCACGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9412
 Oy 421 GGTGTTGGCAACAGCGCGGTAAACGATTTAA 456
 Db 9413 GCGTTGGTAAACAGCGCGGTAAACGATTTAA 9448

FEATURES SOURCE

Takayama, Ikoma, Nara 630-01, Japan
 (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660,
 Fax:81-7437-2-5669)
 Collaboration Information:
 Project:
 The Japan E.coli genome DNA sequencing project
 Group:
 The Japan E.coli genome DNA sequencing group
 Members: (1995.4 - 1996.3)
 Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
 Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
 Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
 Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
 Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
 Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
 Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
 Yamamoto,Y. and Yano,M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kishori@nibb.ac.jp
 Information operator:
 Name: Hirokazu Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmori@gtc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://bsw3.aist-nara.ac.jp/
 location/Qualifiers
 1..15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23..7-24.0 min"
 /clone="Kohara clone #231"
 /note="Nucleotide position 1097101-1112147 from the
 initiation site of *thiA* (0 min.). This clone is from
 Kohara lambda miniset library."
 <1..276
 /note="unnamed protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35822.1"
 /db_xref="GI:4062605"
 /translation="DLFSSSENALSTFFAKPLMLRLFILVIEVMITLIVSLATVE
 WRWFEAYFLICVPLIYEFYCRFLMNDPMACDMYFRMGHFNR"
 1152..2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=1
 /product="Hypothetical 33.0 kD protein in prob-proA
 intergenic region."
 /protein_id="BAA35823.1"
 /db_xref="GI:4062606"
 /translation="MDIIFHPTFDQWIEALRKAIPOARVBAWKSNDSDADVALY
 WHPPEMLAGRDLCAYFALGAVDSILSKLQAPHEMLNPVPLFRLEDTGMEQOMDY
 ALQVLIWFRPFDYRLOONSSHQPLPEYREDPTGIGAGVLSKVAQSLQTFP
 PLRCWSTRSMPGVOSFAGEELSAFLSCRVLIINLPMTPEVGIINOQLLEKLPD
 GAYLIINRGVHVIEDLLAALDSGVKGMALVFNREPLPESSPLMQHRRVITTHV
 AATTPRAVEYISRTIAOLEKGERVCQVDRARGY"
 2145..2882
 /note="unnamed protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=1
 /protein_id="BAA35824.1"
 /db_xref="GI:4062607"
 /translation="MTPVDLHMTVASTAVSTLSYDIAOKGKIKLPAITDHGPM
 EDAPEHMFPTNMRIPRWVGVGILGAEINIKRVGEIDCSQMPSDLIIAGFHS
 PVPAPHDKATNTQAMATIASGVNVIISHGNGRKYBIDYVAVEAAKHQVVALEINS
 SPFHSRKSSEDNCREVAAVRADGAVWALQSDSHSTAFATMGEPFECLEILDAVDFPPER

RESULT 13
 D90741 15047 bp DNA linear BCT 25-DEC-2002
 LOCUS Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
 DEFINITION D90741 AB001340
 ACCSSION D90741.1 GI:1651509
 VERSION D90741.1 GI:1651509
 KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
 csgA; ycdB; cIs; nov; mdog.
 SOURCE Escherichia coli K12
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1
 Oshima,T., Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
 Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
 Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
 Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H.,
 Nishio,Y., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K.,
 Wada,C., Yamamoto,Y., Yano,M. and Horiuchi,T.
 A 718-kb DNA sequence of the Escherichia coli K-12 genome
 corresponding to the 12.7-28.0 min region on the linkage map
 JOURNAL DNA Res. 3 (3), 137-155 (1996)
 MEDLINE 97061202
 PUBMED 8905232
 REFERENCE 2
 Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T.,
 Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K.,
 Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M.,
 Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H.,
 Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T.,
 Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C.,
 Yamamoto,Y. and Yano,M.
 The systematic sequencing of the Escherichia coli genome in Japan
 JOURNAL Unpublished
 REFERENCE 3 (bases 1 to 15047)
 MORI,H.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUL-1996) Hirokazu Mori, NARA Institute of Science
 and Technology, Res. & Edu. Center for Genetic Info.; 8916-5

CDS

```
ILNVSPRLINFLBSRGMAPIAEFADL"
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=1
/product="Hypothetical protein H11543"
/protein_id="BA035825.1"
/db_xref="GI:4062608"
/translation="MNEFSLICRVLGSLYRQPODPLVPLFTLIREKLAANPLEQ
DELTPLQKSCDMTVASADYNALFIDECAPVPSAWEGAETAEVAFTESEGMPL
RTMAPLITGLILAAASWLEDOSTEDESALETLFSBYLLPNCAGALGVKVAHATTPPW
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=1
/protein_id="BA035826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAFSAIRGNMILLSTAITTGILSGIKWVAVSLGL
SVNAGLICTAIVFACPOGQKGLAISATLISGVWMAVITIGSALAHLEILGYITG
LVAFWLCIOAKOLLISFVPGFTIGACATPAQGMKVLPSLALGLISYAMNSGLM
complement(4117..4950)
/gene="cs98"
/complement(4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs98 precursor."
/protein_id="BA035827.1"
/db_xref="GI:4062610"
/translation="MORFLVAVMLISGLTAPPEKAPRTIMPRAQSYKDLTLPLA
NLNARKITIAQENGTVAIINRIPIOSLTANINBESITIGSNVSGGVGATYQ
IGADTQQLDQIVANLVNIVVSTIGTISLVSKITLISBYQAGVFIDYQRLBE
VGTISNEPVM/LCMSAIEFTGVI/FLINDIGDRLMDLQNKARQNDILVKYRHSVPPE
S"
complement(4977..5393)
/gene="cs98"
/complement(4977..5393)
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs98 precursor."
/protein_id="BA035828.1"
/db_xref="GI:4062611"
/translation="MRVKAHVLLMLISPLSNAGTWTFOFRNPNFGNPNNGAFLINS
AOAKSYKDSYNDPFGIETPSALDNFTQALISQILGSLININTGKFERVNDYIV
DIANDQQLDNLVDRITGOTSTIOVSGIQQNSTDP"
complement(5418..5807)
/gene="cs98"
/complement(5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs98 precursor."
/protein_id="BA035829.1"
/db_xref="GI:4062612"
/translation="MKRYLRMTVAAEFLFAAGNLAAVEVPGILTDHTVSSIGHDFY
RAFSKDSYDNGNTINERSARWGSWITTVQDVI/FTFLFPLKRPKEKIVYFAL
IOTBEAKNRQJNLSTGLADDEF"
complement(5812..6462)
/gene="cs98"
```

CDS

```
complement(5812..6462)
/gene="cs98"
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable cs98B operon transcriptional regulatory
protein."
/protein_id="BA035830.1"
/db_xref="GI:4062613"
/translation="MFEVHSIHGHTLLITSSLOATNLLQHKOSIATIGKLNIG
RSLDDISGSIILLDMHEADKLIHWODTISRKNKIKILNLPEDYPRDIENMP
HINGVFYSMEDQRRVNGOGVIRGCYTOGLASLYLTHGNNRVNTESSALLTHE
KEILNKLRIKASNNELRSLFISENTVKHLNLPFKIAVKNRTOAVSMANDNLR"
7217..7672
/gene="cs98B"
7217..7672
```

gene

CDS

Query Match 56.1%; Score 256; DB 1; Length 15047;
Best Local Similarity 72.6%; Pred. No. 1,2e-62;
Matches 331; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY	1	ATGAACTTTTAAAGTGCGACATCTCCAGCAATCGTAGTTCTGCGAGTCTGGCT	60
DB	7713	ATGAACTTTTAAAGTGCGACATCTCCAGCAATCGTAGTTCTGCGAGTCTGGCT	60
QY	61	GGCGTGTTCGCAATGCGCGCGCGGTATCATTAACGCGCGCAATAGTTCCGCG	120
DB	7773	GGGTGTTCCTCAGTACGGCGCGCGGTATCATTAACGCGCGCAATAGTTCCGCG	120
QY	121	CGGACTCAAGTGTGAGCATTTATCAGTACGTTCCGCTAACCGTCCGCTCGCA	180
DB	7833	CCAAATTCGAGCTGACATTTACCACTACGAGTACGAGTACGCTGCACTTCGCA	180
QY	181	AGCAATGCCCTAATCTGAAAGCACCATTACCGAGCGGTATGAGCGCCGCAAT	240
DB	7893	ACTATGCGCTAATCTGACTGACTGACTGACTGACTGACTGACTGACTGACTG	240
QY	241	GTAAGCCAGGCGGTGATTAAGTACTATTAAGTACTGATTAAGTACTGATTAAG	300
DB	7953	GTTGTCAGGCGTCAATGACAGCTCAATGACAGCTCAATGACAGCTCAATGAC	300
QY	301	GCCACATGACAGCGTGAACGCTGAAATCTCGATATTAAGTACTGATTAAGT	360
DB	8013	GCTACTTGTATGATGAGGAAACGCAAAATCTGAAATGACGTTAAACAGTTGG	360
QY	361	AATPAACGCGCGCTGATTAATGACAGCGCATCTGATTAATGACAGCGCAT	420
DB	8073	GCGAACGCGTGTGACAGTGAACAGCTGATTAATGACAGCGCATCTGATTA	420
QY	421	GCTTTGGCAACAGCGCATCTGATTAATGACAGCGCATCTGATTAATGACAG	480
DB	8133	GCTTTGGCAACAGCGCATCTGATTAATGACAGCGCATCTGATTAATGACAG	480

gene

CDS

```
RESULT 14
AX814811
LOCUS AX814811 456 bp DNA linear PAT 05-DEC-2003
DEFINITION Sequence 15 from Patent WO0306446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Bjorck, L., Olsen, A., Wikstrom, M. and Herwald, H.
Peptides
Patent: WO 0306446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
Location/Qualifiers
1..456
```

gene

CDS

/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1. .456
/note="unamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAB83541.1"
/db_xref="GI:39104002"
/translation="MKLLKVEAIAIVFSGSALAGVVPQYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLTIQHGSGNADVGGSDDSSIDLTRGFNSAT
LDQWNGKNSMTVKQFGGNGAAYDQFASNSVNVTVGVFGNATAHQY"

ORIGIN

Query Match 55.8%; Score 254.4; DB 6; Length 456;
Best Local Similarity 72.4%; Pred. No. 2.8e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTCGACGATTCGAGTTCTGGCAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTCGACGATTCGAGTTCTGGCAGTCTTGCT 60
QY 61 GGCCTCGTTCCACATGGGCGCGCGGTAATATACGCGCGGCAATAGTTCGCGC 120
DB 61 GGTGTGTCTCTCACTAGCGCGCGGCGGTAATATACGCGCGGCAATAGTTCGCGC 120
QY 121 CCGACTCAGCGTTGAGATTATCAGCGCTTCGCTTAACGCTGCGCTTCTTGCAA 180
DB 121 CCAAAATCTGAGCTGAACATTACAGTACCGTGGCGGTAATCTGCACTTCTTGCAA 180
QY 181 AGGATGCCGCGTAATATCGAAGACCACTTACGAGCGGTTATGGTAACGGCGCGAT 240
DB 181 ACTATGCCGCGTAATCTGACTTATCTATTAACCGATGCGCGGTAATGGTGCAGAT 240
QY 241 GTAGCCAGGCGTGGGATATATAGTATTAAGTCACTGCAATGGTTTCAGAAATAT 300
DB 241 GTTGGTACGGCTCAGATGACAGCTCATGATCTGACCAAGCGTGGTTCGTAACGC 300
QY 301 GCCACATCGACGATGGAACGCTAAATCTCCGATTTACTGTCGCGCAATACGGCGGT 360
DB 301 GCTACTCTTGATCAGTGGAACGGCAAAATCTGAATGACGCTTAAACATGTCGGTGGT 360
QY 361 AATAAGCCGCGGTGTTATCATACGCGATCTGATTCAGCGTAATGGTCCGTCAGGT 420
DB 361 GGCACGCGTGTGAGTTGACCAAGCTGATCTCTCCGTCACGATCAGGT 420
QY 421 GGTTCGCAACACGCGCGCGCTTAACCACTATTA 456
DB 421 GCGTTGGTAACAACGCGCGCTCATCAGTACTTA 456

RESULT 15
ECOCSSGA 648 bp DNA linear BCT 13-JUL-1993

LOCUS Escherichia coli curlin subunit (csgA) gene, complete cds.
DEFINITION L049379.1 GI:290424

VERSION L049379.1 GI:290424
KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.

SOURCE Escherichia coli
ORGANISM Escherichia coli

REFERENCE 1 (bases 1 to 648)
Olsen, A., Arqvist, A., Hammar, M., Skupolvi, S. and Normark, S.

THE RPOD sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli
in Escherichia coli

Mol. Microbiol. 7 (4), 523-536 (1993)

JOURNAL 93211294

MEDLINE 8459772

COMMENT On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara) DNA.

FEATURES
SOURCE

Location/Qualifiers
1. .648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/c189ue_11b="Kohara"
83. .538
/gene="csgA"
83. .538
/gene="csgA"
/note="precursor protein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="curlin subunit"
/protein_id="AAA23616.1"
/db_xref="GI:290425"
LNIYQYGGNSALALQTDARNSDLTIQHGSGNADVGGSDDSSIDLTRGFNSAT
LDQWNGKNSMTVKQFGGNGAAYDQFASNSVNVTVGVFGNATAHQY"
83. .142
/gene="csgA"
/evidence=experimental
143. .535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

ORIGIN

Query Match 55.8%; Score 254.4; DB 1; Length 648;
Best Local Similarity 72.4%; Pred. No. 2.9e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTCGACGATTCGAGTTCTGGCAGTCTTGCT 60
DB 83 ATGAACTTTTAAAGTGGCAGATTCGACGATTCGAGTTCTGGCAGTCTTGCT 60
QY 61 GGCCTCGTTCCACATGGGCGCGGTAATATACGCGCGGCAATAGTTCGCGC 120
DB 143 GGTGTGTCTCTCACTAGCGCGCGGCGGTAATATACGCGCGGCAATAGTTCGCGC 120
QY 121 CCGACTCAGCGTTGAGATTATCAGCGCTTCGCTTAACGCTGCGCTTCTTGCAA 180
DB 121 CCAAAATCTGAGCTGAACATTACAGTACCGTGGCGGTAATCTTGCACTTCTTGCAA 180
QY 181 AGGATGCCGCGTAATATCGAAGACCACTTACGAGCGGTTATGGTAACGGCGCGAT 240
DB 263 ACTATGCCGCGTAATCTGACTTATCTATTAACCGATGCGCGGTAATGGTGCAGAT 240
QY 241 GTAGCCAGGCGTGGGATATATAGTATTAAGTCACTGCAATGGTTTCAGAAATAT 300
DB 323 GTTGGTACGGCTCAGATGACAGCTCATGATCTGACCAAGTGGTTCGTAACGC 382
QY 301 GCCACATCGACGATGGAACGCTAAATCTCCGATTTACTGTCGCGCAATACGGCGGT 360
DB 383 GCTACTCTTGATCAGTGGAACGGCAAAATCTGAAATGACGCTTAAACAGTTCGGTGGT 442
QY 361 AATAAGCCGCGGTGTTATCATACGCGATCTGATTCAGCGTAATGGTCCGTCAGGT 420
DB 443 GGCACGCGTGTGAGTTGACCAAGCTGATCTTCACTTCTCCGTCACGATCAGGT 502
QY 421 GGTTCGCAACACGCGCGCGCTTAACCACTATTA 456
DB 503 GCGTTGGTAACAACGCGCGCTCATCAGTACTTA 538

Search completed: March 15, 2004, 22:49:37
Job time : 1972.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds

(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 1 atgaacttttaaaagtcgac.....ccacgcctaacagctattaa 456

Sequence: 1 atgaacttttaaaagtcgac.....ccacgcctaacagctattaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: Geneseq1990s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001as:*

5: Geneseq2001bs:*

6: Geneseq2002s:*

7: Geneseq2003as:*

8: Geneseq2003bs:*

9: Geneseq2003cs:*

10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	2	AAQ87467
2	456	100.0	456	2	AAQ87467
3	456	100.0	456	2	AAQ87467
4	456	100.0	456	2	AAQ87467
5	456	100.0	456	2	AAQ87467
6	456	100.0	456	2	AAQ87467
7	456	100.0	456	2	AAQ87467
8	456	100.0	456	2	AAQ87467
9	456	100.0	456	2	AAQ87467
10	456	100.0	456	2	AAQ87467
11	456	100.0	456	2	AAQ87467
12	456	100.0	456	2	AAQ87467
13	456	100.0	456	2	AAQ87467
14	456	100.0	456	2	AAQ87467
15	456	100.0	456	2	AAQ87467
16	456	100.0	456	2	AAQ87467
17	456	100.0	456	2	AAQ87467
18	456	100.0	456	2	AAQ87467
19	456	100.0	456	2	AAQ87467
20	456	100.0	456	2	AAQ87467
21	456	100.0	456	2	AAQ87467
22	456	100.0	456	2	AAQ87467
23	456	100.0	456	2	AAQ87467

24	42.4	9.3	100	7	ACD68808
25	40.8	8.9	100	7	ACD68809
26	35.4	7.8	3411	5	AA888526
27	35.4	7.8	3412	5	AA889144
28	34.6	7.6	456	3	AA664620
29	34	7.5	456	3	AA664618
30	34	7.5	1344	2	AA090007
31	34	7.5	1344	3	AA000669
32	34	7.5	1344	3	AA014939
33	34	7.5	1344	6	AA041133
34	34	7.5	1344	6	AA029125
35	34	7.5	1344	6	AA027016
36	34	7.5	1344	6	AB051711
37	34	7.5	2886	7	ACC59886
38	34	7.5	3300	7	AA255699
39	33.4	7.3	3990	2	AA061785
40	33.4	7.3	5100	2	AA020271
41	33.4	7.3	7766	2	AA085043
42	33.2	7.3	788	4	AA080100
43	32.8	7.2	78	3	AA064609
44	32.8	7.2	1005	7	ACA02133
45	32.8	7.2	1059	5	AA065447

ALIGNMENTS

RESULT 1	AAQ87467	standard; DNA; 456 bp.
ID	AAQ87467	
XX	AAQ87467	
AC	AAQ87467	
XX	25-MAR-2003	(revised)
DT	26-JUN-1995	(first entry)
XX	Agfa sequence.	
DE	Salmonella; Agfa; vaccine; genetic immunization; ds.	
XX	Salmonella.	
XX	Salmonella.	
OS	Salmonella.	
XX	Key	Location/Qualifiers
FT	CDS	1..454
FT		/*tag= a
FT		/note= "Agfa"
XX	MO9425598-A2.	
XX	10-NOV-1994.	
PD	26-APR-1994;	94WO-IB000207.
XX	26-APR-1993;	93US-00054452.
PR	(UVI-) UNIV VICTORIA INNOVATION & DEV CORP.	
XX	(KING/) KING J.	
PA	Kay WW, Collinson SK, Clouthier SC, Doran JL;	
PI	WPI; 1994-358275/44.	
XX	P-PSDB; AAR74625.	
DR	Eliciting an immune response to Salmonella - using attenuated Salmonella	
XX	strains, vector constructs, or compans. contg. fimbrial type proteins.	
PT	Disclosure; Fig 7B; 95pp; English.	
XX	The DNA encodes the Salmonella Agfa protein. The DNA and isolated	
CC	proteins are used in genetic immunization and vaccine compositions, (e.g.	
CC	respectively, to elicit an immune response to Salmonella in animals (e.g.	
CC	food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN	
CC	field.)	

XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 456; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-118;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGCAGCATTCGACGCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
DB 61 GGGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
QY 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACGCTGGCTTGGCTGGCA 180
DB 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACGCTGGCTTGGCTGGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGATCTATTGAATCTGACATGATGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGGATTAATGATCTATTGAATCTGACATGATGTTTCAGAAATAT 300
QY 301 GCCACCATCGACAGTGGAAAGCTAAACTCCGATTAATCTGTGGCCAAATACGGCGGT 360
DB 301 GCCACCATCGACAGTGGAAAGCTAAACTCCGATTAATCTGTGGCCAAATACGGCGGT 360
QY 361 AATAAGCCCGCGTGGTATTCAGACCGCATCTGATTCAGCGTATAGTGGTCAAGTT 420
DB 361 AATAAGCCCGCGTGGTATTCAGACCGCATCTGATTCAGCGTATAGTGGTCAAGTT 420
QY 421 GGTTTGGCAACACCGCAACGCTAACCGATTAATA 456
DB 421 GGTTTGGCAACACCGCAACGCTAACCGATTAATA 456

RESULT 2

AAT74142
ID AAT74142 standard; DNA; 456 BP.
AC AAT74142;
XX

DT 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX

DE Salmonella enteritidis 27655-3b agfa gene.
XX

KW Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
OS Salmonella enteritidis.
XX

Key Location/Qualifiers
FT CDS 1..456
FT /tag= a
FT /label= agfa_gene_fragment
FT /transl_except= (pos:367..369,aa:Pro)
XX

US5635617-A.
XX

03-JUN-1997.
XX

26-APR-1994; 94US-00233788.
XX

26-APR-1993; 93US-00054452.
XX

(UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX

Collinson SK, Kay WW, Doran JL;
XX

DR WPI; 1997-30986/28.
XX
P-PSDB; AAM23570.
XX

PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
PS Claim 1; Col 19-112; 85bp; English.
XX

CC The present sequence represents an isolated agfa gene derived from
CC Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC family Enterobacteria. It can also be used to provide proteins and
CC antibodies which can be used for assays. The nucleic acid sequence can be
CC used to provide probes or primers which can specifically hybridize to
CC nucleic acid molecules from greater than 99% of Salmonella strains that
CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC correct PF field.)
XX

SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
Query Match 100.0%; Score 456; DB 2; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.2e-118;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGCAATCGTAGTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGCAGCATTCGACGCAATCGTAGTTCTGGCAGTGTCTGGCT 60
QY 61 GGGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
DB 61 GGGGTGTTCCAAATGGGGCGGGCGGGTATCATTAACGGCGCGCATAGTTCGGC 120
QY 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACGCTGGCTTGGCTGGCA 180
DB 121 CCGGACTCAAGCTTGAAGTATTCAGTACGTTCCGCTAACGCTGGCTTGGCTGGCA 180
QY 181 AGCGATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCCGTAATCTGAACGACCATTAACCGAGCGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGATCTATTGAATCTGACATGATGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGGATTAATGATCTATTGAATCTGACATGATGTTTCAGAAATAT 300
QY 301 GCCACCATCGACAGTGGAAAGCTAAACTCCGATTAATCTGTGGCCAAATACGGCGGT 360
DB 301 GCCACCATCGACAGTGGAAAGCTAAACTCCGATTAATCTGTGGCCAAATACGGCGGT 360
QY 361 AATAAGCCCGCGTGGTATTCAGACCGCATCTGATTCAGCGTATAGTGGTCAAGTT 420
DB 361 AATAAGCCCGCGTGGTATTCAGACCGCATCTGATTCAGCGTATAGTGGTCAAGTT 420
QY 421 GGTTTGGCAACACCGCAACGCTAACCGATTAATA 456
DB 421 GGTTTGGCAACACCGCAACGCTAACCGATTAATA 456

RESULT 3

AAC64617
ID AAC64617 standard; DNA; 456 BP.
AC AAC64617;
XX

DT 26-FEB-2001 (first entry)
XX

DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
XX

KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX

OS Salmonella enteritidis.
XX

PN MO20060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36341.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 134; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
XX assembly system of strains of *Salmonella*, *Escherichia coli* and
XX *Enterobacteriaceae* for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant AgfA
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a *Salmonella*, *E. coli* or
XX *Enterobacteriaceae* host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention
XX
XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 456; DB 3; Length 456;
XX Best Local Similarity 100.0%; Pred. No. 8.2e-118;
XX Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGAACCTTTAAAGTGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 60
XX 1 ATGAACCTTTAAAGTGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 60
XX 1 ATGAACCTTTAAAGTGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 60
XX 61 GCGCGTCTTCCACATATGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 120
XX 61 GCGCGTCTTCCACATATGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 120
XX 61 GCGCGTCTTCCACATATGGGCGGCGGCGGTATCATTAACGGCGGCGGCAATAGTTCCGGC 120
XX 121 CCGGACTTAACGTTGAGCATTTATAGTACGTTCCGCTAACGCTGCGCTGCTCTGCA 180
XX 121 CCGGACTTAACGTTGAGCATTTATAGTACGTTCCGCTAACGCTGCGCTGCTCTGCA 180
XX 121 CCGGACTTAACGTTGAGCATTTATAGTACGTTCCGCTAACGCTGCGCTGCTCTGCA 180
XX 121 CCGGACTTAACGTTGAGCATTTATAGTACGTTCCGCTAACGCTGCGCTGCTCTGCA 180
XX 181 AGCGATGCGCGGTAATCGAATCGATTAACCGAGCGGTTATGTAACGGCGCGCAT 240
XX 181 AGCGATGCGCGGTAATCGAATCGATTAACCGAGCGGTTATGTAACGGCGCGCAT 240
XX 181 AGCGATGCGCGGTAATCGAATCGATTAACCGAGCGGTTATGTAACGGCGCGCAT 240
XX 241 GTAGGCCAGGGTGGCGATTAATAGTACTTATGAACGACTCAGATGTTTCAGAAATAT 300
XX 241 GTAGGCCAGGGTGGCGATTAATAGTACTTATGAACGACTCAGATGTTTCAGAAATAT 300
XX 241 GTAGGCCAGGGTGGCGATTAATAGTACTTATGAACGACTCAGATGTTTCAGAAATAT 300

XX 301 GCCACATTCGACCACTGGTAACGCTTAAABAACTCCGATATTACTGCGGCCAATACGCGGT 360
XX 301 GCCACATTCGACCACTGGTAACGCTTAAABAACTCCGATATTACTGCGGCCAATACGCGGT 360
XX 361 AATAACGCGCGCGTGGTTAATCAGACCGCATCTGATTCACCGTAATAGTGCAGGTT 420
XX 361 AATAACGCGCGCGTGGTTAATCAGACCGCATCTGATTCACCGTAATAGTGCAGGTT 420
XX 421 GGTTCGCAACAAACGCGCGGCTTAACCGATTTAA 456
XX 421 GGTTCGCAACAAACGCGCGGCTTAACCGATTTAA 456
XX
XX RESULT 4
XX AAC64626
XX ID AAC64626 standard; DNA; 456 BP.
XX AAC64626;
XX 26-FEB-2001 (first entry)
XX
XX AgfA::PT3#5 DNA sequence SEQ ID NO:19.
XX
XX *Salmonella*; agfA; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX *Salmonella enteritidis*.
XX *Escherichia coli*.
XX Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36350.
XX
XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
XX assembly system of strains of *Salmonella*, *Escherichia coli* and
XX *Enterobacteriaceae* for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbrial subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant AgfA
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a *Salmonella*, *E. coli* or
XX *Enterobacteriaceae* host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant AgfA protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
 Best Local Similarity 94.1%; Pred. No. 9, 2e-124;
 Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTTGCT 60
QY 61 GGGGTGTTCCCAATGGGCGCGCGCGGTATCTAATACGGCGGCAATAGTTCGGC 120
DB 61 GGGGTGTTCCCAATGGGCGCGCGCGGTATCTAATACGGCGGCAATAGTTCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCCAGACGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCCAGACGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGCGCAGGTCGGGATATAGTACTATGAACTGATGATGATGATGATGATGATGAT 300
DB 241 GTAGCGCAGGTCGGGATATAGTACTATGAACTGATGATGATGATGATGATGATGAT 300
QY 301 GCCACATTCGACCGTGAACGCTAAACTCGATATTAAGTGTGGCAATAGCGCGGT 360
DB 301 GCCACATTCGACCGTGAACGCTAAACTCGATATTAAGTGTGGCAATAGCGCGGT 360
QY 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGGTTATGTCGTCAGGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGGTTATGTCGTCAGGTT 420
QY 421 GGTTCGGCAACACCGCGCTAACGATATTA 456
DB 421 GGTTCGGCAACACCGCGCTAACGATATTA 456

```

RESULT 5

AAC64625 ID AAC64625 standard; DNA; 456 BP.

AC AAC64625;
 XX
 DT 26-FEB-2001 (first entry)
 XX

DE Agfa: PT3#4 DNA sequence SEQ ID NO:17.

XX Salmonella agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.

XX WO200060102-A2.
 XX 12-OCT-2000.
 XX

XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX

XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay WW;
 PT

XX
 DR WPI: 2000-672631/65.
 DR P-PSDB; AAB36349.
 XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PS protein useful for eliciting immune response in animal.
 XX Disclosure, Page 136, 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7A) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
 Best Local Similarity 93.2%; Pred. No. 1.1e-121;
 Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTTGCT 60
QY 61 GGGGTGTTCCCAATGGGCGCGCGCGGTATCTAATACGGCGGCAATAGTTCGGC 120
DB 61 GGGGTGTTCCCAATGGGCGCGCGCGGTATCTAATACGGCGGCAATAGTTCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGCTTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCCAGACGGTTATGGTAAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCCAGACGGTTATGGTAAACGGCGCAT 240
QY 241 GTAGCGCAGGTCGGGATATAGTACTATGAACTGATGATGATGATGATGATGATGAT 300
DB 241 GTAGCGCAGGTCGGGATATAGTACTATGAACTGATGATGATGATGATGATGATGAT 300
QY 301 GCCACATTCGACCGTGAACGCTAAACTCGATATTAAGTGTGGCAATAGCGCGGT 360
DB 301 GCCACATTCGACCGTGAACGCTAAACTCGATATTAAGTGTGGCAATAGCGCGGT 360
QY 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGGTTATGTCGTCAGGTT 420
DB 361 AATAACCGCGCGTGTATATCAGACCGCATCGATTCAGGTTATGTCGTCAGGTT 420
QY 421 GGTTCGGCAACACCGCGCTAACGATATTA 456
DB 421 GGTTCGGCAACACCGCGCTAACGATATTA 456

```


Db 421 GGTTTGGCAACAGCCAGGCTAACGATATTAA 456

RESULT 6
AAC64628 standard; DNA; 456 BP.

XX ID AAC64628
XX AC AAC64628;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX KM vaccine; immune response; immunogen; ds.
XX OS Salmoneilla enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN MO20060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000MO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UYVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX DR WPI; 2000-672631/65.
XX DR P-PSDB; AAB36352.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEFI/TAI) nucleation depended assembly system of strains of Salmoneilla, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmoneilla, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
Best Local Similarity 93.0%; Pred. No. 3.7e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCGCATTTGGCAGCAATCGTAGTTTCTGGCATGCTGCT 60
Db 1 ATGAAACTTTTAAAGTGGCGCATTTGGCAGCAATCGTAGTTTCTGGCATGCTGCT 60
QY 61 GGGCTCGTTCCACAATGGGGGCGGGGGGTAATCAATAGCGGGCGCATAGTTCCGGC 120
Db 61 GGGCTCGTTCCACAATGGGGGCGGGGGGTAATCAATAGCGGGCGCATAGTTCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGTCGCTTCTGCA 180
Db 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGTCGCTTCTGCA 180
QY 181 AGCGATCGCCGTAATCTGAAACGACCAATTAACCAAGCGGTTATGTAACGGCGCAT 240
Db 181 AGCGATCGCCGTAATCTGAAACGACCAATTAACCAAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGGATTAATAGTACTATTGACTGACTGAGAAAGGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGGGATTAATAGTACTATTGACTGACTGAGAAAGGTTTCAAAATAT 300
QY 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTACTGCGCCATACGGCGGT 360
Db 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTACTGCGCCATACGGCGGT 360
QY 361 AATAACCGCGCGCTGTTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db 361 AATAACCGCGCGCTGTTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTTCGCAACAGCGCAGCGCTAACGATATTAA 456
Db 421 GGTTCGCAACAGCGCAGCGCTAACGATATTAA 456

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.

XX AC AAC64622;
XX DT 26-FEB-2001 (first entry)
XX DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
XX KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX KM vaccine; immune response; immunogen; ds.
XX OS Salmoneilla enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN MO20060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000MO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UYVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX DR WPI; 2000-672631/65.
XX DR P-PSDB; AAB36346.
XX PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX PS Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 121 A, 112 C, 118 G, 105 T, 0 U, 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
 Best Local Similarity 93.0%; Pred. No. 3.7e-121;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
QY 61 GGGCTGCTTCCACAAATGGGGCGGGGGGTAATCAATACGGCGGCAATGTTCCGGC 120
DB 61 GGGCTGCTTCCACAAATGGGGCGGGGGGTAATCAATACGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGACATTTATCATGATCGGTTCCGCTGCAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGACATTTATCATGATCGGTTCCGCTGCAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAATCTGAAGACCACTTACCCAGACGGTTATGTTACGCGCGAT 240
DB 181 AGCGATGCCCGTAATCTGAAGACCACTTACCCAGACGGTTATGTTACGCGCGAT 240
QY 241 GTAGGCGAGGGTGGCGATTAATGTAATGTAACGTAAGTGTTCAGAAATTAAT 300
DB 241 GTAGGCGAGGGTGGCGATTAATGTAATGTAACGTAAGTGTTCAGAAATTAAT 300
QY 301 GGCACATCGACAGTGGAAAGCGTAAATCTCGATTTTCTGCGCCAAATCGCGCGT 360
DB 301 GGCACATCGACAGTGGAAAGCGTAAATCTCGATTTTCTGCGCCAAATCGCGCGT 360
QY 361 AATAAGCGCGGGTGTATATCAAGCGCATCTGATTCAGCGTAAATGCGTCAGGT 420
DB 361 AATAAGCGCGGGTGTATATATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GATTGGCAACAGCGACGCGCTAACCACTATTA 456
DB 421 GCACATGCAACAGCGACGCGCTAACCACTATTA 456

```

RESULT 8
 AAC64629 standard; DNA; 456 BP.

AC AAC64629;
 DT 26-FEB-2001 (first entry)
 XX

DE Agfa::PT3#8 DNA sequence SEQ ID NO:25.
 XX *Salmonella*: agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX WO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UUVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay W;
 PI WPI: 2000-672631/65.
 DR P-PSDB; AAB36353.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF7/7AF) nucleation depended
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombinant gene into the chromosome of the homologous species;
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP, 114 A, 108 C, 123 G, 111 T, 0 U, 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;
 Best Local Similarity 93.0%; Pred. No. 3.7e-121;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACGATCGTGTGCTGCTGCT 60
QY 61 GGGCTGCTTCCACAAATGGGGCGGGGGGTAATCAATACGGCGGCAATGTTCCGGC 120
DB 61 GGGCTGCTTCCACAAATGGGGCGGGGGGTAATCAATACGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGACATTTATCATGATCGGTTCCGCTGCAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGACATTTATCATGATCGGTTCCGCTGCAACGCTGCTGCTGCA 180

```

DB 121 CCGAGCTCAAGCTTGAGCATTTATCATGACGCTTCGCGTAACGCTGCTTTATGATCAG 180
QY 101 AGCGATGCCCTGTAATTCGAAACGACCATTCACGAGCGGTTATGTTACGCGCCGAT 240
DB 101 CTGGTTACCCGCTGTTGTTATCCCATGAAATGCGACATGCGGTTATGTTACGCGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATATATGTACTATTTGAACTGACTCAGATAGTGGTTTCGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATATGTACTATTTGAACTGACTCAGATAGTGGTTTCGAAATAT 300
QY 301 GCCACCATCGACACAGTGGAACCGCTAAACCTCCGATATTACTGTGCGCCAAATACGCGCGT 360
DB 301 GCCACCATCGACACAGTGGAACCGCTAAACCTCCGATATTACTGTGCGCCAAATACGCGCGT 360
QY 361 AATAACGCGCGCTGCTGTTATATGACACCGCATTTGATTCAGCGTATGTCGCTAGGTT 420
DB 361 AATAACGCGCGCTGCTGTTATATGACACCGCATTTGATTCAGCGTATGTCGCTAGGTT 420
QY 421 GGTTCGCAACACGCGCGCTAACCATGATTAA 456
DB 421 GGTTCGCAACACGCGCGCTAACCATGATTAA 456

RESULT 9
AAC64623

ID AAC64623 standard; DNA; 456 BP.

AAC64623;

26-FEB-2001 (first entry)

Agfa::PT3#2 DNA sequence SEQ ID NO:13.

Salmone11a; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmone11a enteritidis.

Escherichia coli.

Synthetic.

MO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UVVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay MW;

WPI; 2000-672631/65.

P-PSDB; AAB36347.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.

Discloure; Page 136; 139p; English.

CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/TAF) nucleation depended
CC assembly system of strains of Salmone11a, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae competitively;
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively;
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmone11a, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 3; Length 456;

Best Local Similarity 93.0%; Pred. No. 3.7e-121;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACATTCGACAGCAATCGATGTTCTGCGAGTCTTGCT 60

DB 1 ATGAAACTTTTAAAGTGGACATTCGACAGCAATCGATGTTCTGCGAGTCTTGCT 60

QY 61 GCGCTGTTCCACATGCGGCGCGCGGCTATATCAATGCGCGCGCAATATGTTCCGCG 120

DB 61 GCGCTGTTCCACATGCGGCGCGCGGCTATATCAATGCGCGCGCAATATGTTCCGCG 120

QY 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGAGTTCCGCTAACGCTGCTGCTGCA 180

DB 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGAGTTCCGCTAACGCTGCTGCTGCA 180

QY 181 AGCGATGCCCTGTAATTCGAAACGACCATTCACGAGCGGTTATGTTACGCGCCGAT 240

DB 181 AGCGATGCCCTGTAATTCGAAACGACCATTCACGAGCGGTTATGTTACGCGCCGAT 240

QY 241 GTAGGCCAGGGTGGCGATATATGTACTATTTGAACTGACTCAGATAGTGGTTTCGAAATAT 300

DB 241 GTAGGCCAGGGTGGCGATATATGTACTATTTGAACTGACTCAGATAGTGGTTTCGAAATAT 300

QY 301 GCCACCATCGACACAGTGGAACCGCTAAACCTCCGATATTACTGTGCGCCAAATACGCGCGT 360

DB 301 GCCACCATCGACACAGTGGAACCGCTAAACCTCCGATATTACTGTGCGCCAAATACGCGCGT 360

QY 361 AATAACGCGCGCTGCTGTTATATGACACCGCATTTGATTCAGCGTATGTCGCTAGGTT 420

DB 361 CTGGTTACCCGCTGTTGTTATCCCATGAAATGCGACATGCAACGCTATATGTCGCTAGGTT 420

QY 421 GGTTCGCAACACGCGCGCTAACCATGATTAA 456

DB 421 GGTTCGCAACACGCGCGCTAACCATGATTAA 456

RESULT 10
AAC64624

ID AAC64624 standard; DNA; 456 BP.

AAC64624;

26-FEB-2001 (first entry)

Agfa::PT3#3 DNA sequence SEQ ID NO:15.

Salmone11a; agfa; chromosomal gene replacement; fimbria; epitope;
vaccine; immune response; immunogen; ds.

Salmone11a enteritidis.

Escherichia coli.

Synthetic.

MO200060102-A2.

CC Against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;

Best Local Similarity 92.5%; Pred. No. 4.1e-120;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT

1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT

61 GGCCTGTTCCACAAATGGGGGGGGGGGGTATCATTAACGGCGCGCAATAGTTCCGGC

61 GGCCTGTTCCACAAATGGGGGGGGGGGGTATCATTAACGGCGCGCAATAGTTCCGGC

121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGGTACGGCTGCTCTGCA

121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCGGTACGGCTGCTCTGCA

181 AGCGATGCGCGTAAATCTGAACGACCATTAACGAGCGGTTATGTAACGGCGCGAT

181 AGCGATGCGCGTAAATCTGAACGACCATTAACGAGCGGTTATGTAACGGCGCGAT

241 GTAGCCAGGTCGGATTAATGTAATCTTGAATGTAATGTAATGTAATGTAAT

241 GTAGCCAGGTCGGATTAATGTAATCTTGAATGTAATGTAATGTAATGTAAT

301 GCCACATCGACGATGGAACGTAATAAATCCGATTTATCTGCGGCAATACGGCGGT

301 GCCACATCGACGATGGAACGTAATAAATCCGATTTATCTGCGGCAATACGGCGGT

361 AATTAAGCGCGGCTGTTATCAGACCGCATCTGATTCAGCGTATGTCGTCAGTT

361 AATTAAGCGCGGCTGTTATCAGACCGCATCTGATTCAGCGTATGTCGTCAGTT

421 GGTGGGCAACACGCGCGGCTTAACGATTTAA 456

421 GGTGGGCAACACGCGCGGCTTAACGATTTAA 456

RESULT 12

AAC64627 standard; DNA; 456 BP.

AAC64627;

26-FEB-2001 (first entry)

Agfa::PT3#6 DNA sequence SEQ ID NO:21.

Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;

Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;

Salmonella enteritidis.

Becherichia coli.

Synthetic.

MO20060102-A2.

12-OCT-2000.

05-APR-2000; 2000MO-CA000356.

05-APR-1999; 99US-0127888P.

(UVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay WW;

WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SFA/TF) nucleation depended

CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a Salmonella, E. coli or

CC Enterobacteriaceae host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (1) is

CC useful for the expression of recombinant Agfa protein which is useful for

CC eliciting an immune response in an animal. In a fimbrial presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fimbria protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 88.1%; Score 401.6; DB 3; Length 456;

Best Local Similarity 92.5%; Pred. No. 4.1e-120;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT

1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGCGAGTCTTGCT

61 GGCCTGTTCCACAAATGGGGGGGGGGGGTATCATTAACGGCGCGCAATAGTTCCGGC

61 GGCCTGTTCCACAAATGGGGGGGGGGGGTATCATTAACGGCGCGCAATAGTTCCGGC

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT#10 DNA sequence SEQ ID NO:29.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN MO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UUVI-) UNIV VICTORIA.
XX
FI White AP, Doran JL, Collison SK, Kay MW;
XX
DR WPI; 2000-672631/65.
XX
DR P-PSDB; AAB36355.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 87.4%; Score 398.4; DB 3; Length 456;
Best Local Similarity 92.1%; Pred. No. 4.5e-119;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

1 ATGAACCTTTAAAGTGGCAGCATTCGACGATCGATGTTCTTGGCAGTCTTGCT 60
XX
XX

Db 1 ATGAACCTTTAAAGTGGCAGCATTCGACGATCGATGTTCTTGGCAGTCTTGCT 60
QY 61 GGCCTGCTTCAACAATGGGGGGGGCGGCTAATCAATACGGCGGCAATAGTCCGCG 120
Db 61 GGCCTGCTTCAACAATGGGGGGGGCGGCTAATCAATACGGCGGCAATAGTCCGCG 120
QY 121 CCGAATCTCAAGCTTGAACATTTATCAGTACGCTTCGCTTAAGCTGCTGCTGCA 180
Db 121 CCGAATCTCAAGCTTGAACATTTATCAGTACGCTTCGCTTAAGCTGCTGCTGCA 180
QY 181 AGCGATCCCGGTAATATGTAAGACCATTTACCCAGAGCGTTATGTAACGCGCCGAT 240
Db 181 AGCGATCCCGGTAATATGTAAGACCATTTACCCAGAGCGTTATGTAACGCGCCGAT 240
QY 241 GTAGCCAGAGGTGCGGATTAATAGTACTATGTAAGTGAAGTGTTCAGAAATAT 300
Db 241 GTAGCCAGAGGTGCGGATTAATAGTACTATGTAAGTGAAGTGTTCAGAAATAT 300
QY 301 GCCACCATGACCAAGTGGAAACGCTAATAAATCCGATATTACTGTGCGCAATAACGCGCGT 360
Db 301 GCCACCATGATGATGCTGTGTTACCCGATGTAATGACATGACAGCGCGT 360
QY 361 AATTAACCGCGCGCTGTTATCAGACCGCATGATTCACCGTATGTCCTCAGATT 420
Db 361 AATTAACCGCGCGCTGTTATCAGACCGCATGATTCACCGTATGTCCTCAGATT 420
QY 421 GGTTTGGGAACAAGCCACGCGCTTAACCAAGTATTAA 456
Db 421 GGTTTGGGAACAAGCCACGCGCTTAACCAAGTATTAA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
XX AAQ73066;
XX
AC 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
DE
XX
XX Salmonella; Agfa; vaccine; genetic immunization; ds.
KM
XX
OS Salmonella enteritidis.
CC
XX
FH Key Location/Qualifiers
FT 1..359
FT /*tag= a
FT /note= "Agfa"
FT /*tag= b
FT 37..60
FT /*tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT 52..69
FT /*tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT complement(103..129)
FT /*tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT complement(292..402)
FT /*tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
XX MO9425598-A2.
PN
XX
PD 10-NOV-1994.
XX
XX 26-APR-1994; 94WO-IB000207.
PR 26-APR-1993; 93US-00054452.
XX

PA (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
P1 Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX MPI: 1994-358275/44.
XX P-PSDB; AAR62761.
XX
XX Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. confg. fimbrial type proteins.
XX
XX Disclosure: Fig 7A; 95pp; English.
XX
XX The DNA encodes the Salmonella enteritidis 27655-3b TnpHox mutant strain
CC agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
XX Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 331.8; DB 2; Length 361;
Best Local Similarity 99.4%; Pred. No. 1.9e-97;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 64 GTCGTTCCACATGCGGCGCGCGGCGGATATCATTAACGCGCGGCAATAGTTCCGCCCG 123
DB 1 GTCGTACACACAGTGGCGCGCGCGGATATCATTAACGCGCGGCAATAGTTCCGCCCG 60
QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCGCTTGTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCGCTTGTCTGCAAGC 120
QY 184 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCATGTA 243
DB 121 GATGCGCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCATGTA 180
QY 244 GGCAGGCGTGGATTAATGTAATCTTGAATGACTGCAATGTTTCAAGAAATATGCC 303
DB 181 GGCAGGCGTGGATTAATGTAATCTTGAATGACTGCAATGTTTCAAGAAATATGCC 240
QY 304 ACCATGACGAGTGAACGCTAAACAACTCCGATATTATGTCGGCCAAATACGGCGTAT 363
DB 241 ACCATGACGAGTGAACGCTAAACAACTCCGATATTATGTCGGCCAAATACGGCGTAT 300
QY 364 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 335
RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
XX AAT74141;
XX AC
XX 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
XX Salmonella enteritidis 27655-3b TnpHox mutant agfa gene fragment.
XX
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
XX Salmonella enteritidis.
XX
XX Key location/Qualifiers
XX CDS 1..360
FT /*tag= a
FT /*label= agfa_gene_fragment
FT primer_bind 16..60
FT /*tag= b
FT /*label= Primer_TAF5

FT primer_bind 52..69
FT /*tag= c
FT /*label= Primer_TAF3
FT primer_bind complement(103..128)
FT /*tag= d
FT /*label= Primer_TAF6
FT primer_bind complement(1294..312)
FT /*tag= e
FT /*label= Primer_TAF4
PN
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX
XX MPI: 1997-309886/28.
XX
XX P-PSDB; AAM23569.
XX
XX Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
PT enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated agfa gene fragment derived
CC from Salmonella enteritidis 27655-3b TnpHox mutant strain. The nucleic
CC acid can be used to provide diagnostic assays for Salmonella and/or
CC enteropathogenic bacteria of the family Enterobacteria. It can also be
CC used to provide proteins and antibodies which can be used for assays. The
CC nucleic acid sequence can be used to provide probes or primers which can
CC specifically hybridise to nucleic acid molecules from greater than 98% of
CC Salmonella strains that are pathogenic to warm-blooded animals relative
CC to nucleic acid molecules from virtually all other microbial organisms.
CC (Updated on 25-MAR-2003 to correct PF field.)
XX
XX Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
SQ
Query Match 72.8%; Score 331.8; DB 2; Length 361;
Best Local Similarity 99.4%; Pred. No. 1.9e-97;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 64 GTCGTTCCACATGCGGCGCGCGGCGGATATCATTAACGCGCGGCAATAGTTCCGCCCG 123
DB 1 GTCGTACACACAGTGGCGCGCGCGGATATCATTAACGCGCGGCAATAGTTCCGCCCG 60
QY 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCGCTTGTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAAGCGCTTGTCTGCAAGC 120
QY 184 GATCCCGGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCATGTA 243
DB 121 GATGCGCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTAACGCGCGCATGTA 180
QY 244 GGCAGGCGTGGATTAATGTAATCTTGAATGACTGCAATGTTTCAAGAAATATGCC 303
DB 181 GGCAGGCGTGGATTAATGTAATCTTGAATGACTGCAATGTTTCAAGAAATATGCC 240
QY 304 ACCATGACGAGTGAACGCTAAACAACTCCGATATTATGTCGGCCAAATACGGCGTAT 363
DB 241 ACCATGACGAGTGAACGCTAAACAACTCCGATATTATGTCGGCCAAATACGGCGTAT 300
QY 364 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 398
DB 301 AACGCCGCGCTGTTAATCAGACCGCATCTGATTC 335

Search completed: March 15, 2004, 17:51:39

Thu Mar 18 12:27:54 2004

Job time : 254.684 secs

us-09-543-407-1.rng

Page 12

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456

Sequence: 1 atgaacctttaaagtgc.....ccacgctaccgattataa 456

Scoring table: IDENTITY NUC

Gap 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	1	US-08-233-788A-58
2	331.8	72.8	361	1	US-08-233-788A-56
3	34	7.5	1344	3	US-09-120-927-1
4	34	7.5	1344	4	US-09-431-614-5
5	33.4	7.3	7766	4	US-09-125-619-3
6	31.2	6.8	1491	6	5486473-3
7	31.2	6.8	2436	4	US-09-540-236-492
8	31.2	6.8	65792	4	US-09-596-002-31
9	30.8	6.8	2547	3	US-08-508-761B-1
10	30.8	6.8	1664976	4	US-08-916-421B-1
11	30.6	6.7	4403765	3	US-09-103-840A-2
12	30.6	6.7	4411529	3	US-09-103-840A-1
13	29.8	6.5	1008	4	US-09-252-991A-10904
14	29.8	6.5	1077	4	US-09-252-991A-10818
15	29.8	6.5	1521	4	US-09-252-991A-10865
16	29.6	6.5	422	4	US-09-328-352-2055
17	29.4	6.4	1252	4	US-09-252-991A-7641
18	29.4	6.4	1252	4	US-09-561-756-29
19	29.4	6.4	1252	4	US-09-227-721-29
20	29.4	6.4	1252	4	US-09-954-697-29
21	29.4	6.4	1939	4	US-08-961-527-310
22	29.4	6.4	3494	3	US-09-139-802-200
23	29.4	6.4	3494	4	US-09-659-786-200
24	29.4	6.4	3494	4	US-09-023-655-1006
25	29.4	6.4	3660	4	US-09-252-991A-12269
26	29.4	6.4	536165	4	US-09-214-808-1
27	29.2	6.4	2052	1	US-08-443-104-5

C 28	29.2	6.4	2052	1	US-08-238-130-5	Sequence 6, Appli
C 29	29.2	6.4	2052	1	US-08-442-859-5	Sequence 5, Appli
C 30	29.2	6.4	2052	2	US-08-398-489-5	Sequence 5, Appli
C 31	29.2	6.4	2052	2	US-08-894-772-1	Sequence 1, Appli
C 32	29.2	6.4	2052	2	US-09-207-844-1	Sequence 1, Appli
C 33	29.2	6.4	2052	4	US-09-252-509-1	Sequence 1, Appli
C 34	29.2	6.4	2052	5	PCT-US95-0534-5	Sequence 5, Appli
C 35	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
C 36	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
C 37	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
C 38	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
C 39	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
C 40	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
C 41	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
C 42	28.8	6.3	1662	1	US-08-565-386-2	Sequence 2, Appli
C 43	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
C 44	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
C 45	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1

US-08-233-788A-58

Sequence 58, Application US/08233788A

Patent No. 5635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.

APPLICANT: Kay, William W.

APPLICANT: Collinson, Karen S.

APPLICANT: Clouthier, Sharon C.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

TITLE OF INVENTION: OF SALMONELLA

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,788A

FILING DATE: 26-APR-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C2

TELEPHONE: (206) 622-4900

TELECOMMUNICATION INFORMATION:

TELEFAX: (206) 682-6031

TELETYPE: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:

LENGTH: 456 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

FEATURE:

NAME/KEY: CDS

LOCATION: 1..456

US-08-233-788A-58

Query Match 100.0%; Score 456; DB 1; Length 456;

Best Local Similarity 100.0%; Pred. No. 3e-149;

Matches 456; Conservative 0; Mismatches 0; Indels 0;

QY 1 ATGAACTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGGAGTCTCTGAGCT 60
Db 1 ATGAACTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGGAGTCTCTGAGCT 60
QY 61 GGGGTGCTTCCAAATGAGGCGCGGCGGTAAATCAATACGGGCGGCAATAGTTCCGGC 120
Db 61 GGGGTGCTTCCAAATGAGGCGCGGCGGTAAATCAATACGGGCGGCAATAGTTCCGGC 120
QY 121 CCGAGCTCAACGTTGAGCATTTATCAATACGGTCCGCTAAACGCTGCGCTCTGCA 180
Db 121 CCGAGCTCAACGTTGAGCATTTATCAATACGGTCCGCTAAACGCTGCGCTCTGCA 180
QY 181 AGCGATGCGCGTAAATCTGAAACGACATTTACCGAGCGGCTTATGTTACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACATTTACCGAGCGGCTTATGTTACGGCGCGAT 240
QY 241 GTAGGCGAGGCGGTGAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCGAGGCGGTGAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATCGACCAAGTGAAGCTAAATCTCCGATTTATCTGCGGCGCAATACGGCGGT 360
Db 301 GCCACCATCGACCAAGTGAAGCTAAATCTCCGATTTATCTGCGGCGCAATACGGCGGT 360
QY 361 AATACGCGCGCGGTGTTATCAACCGCATCTGATTCAGCGCTAATGTTGCGTACGGT 420
Db 361 AATACGCGCGCGGTGTTATCAACCGCATCTGATTCAGCGCTAATGTTGCGTACGGT 420
QY 421 GGTGTTGGCAACACGCGCAAGCTAATACAGTATTTAA 456
Db 421 GGTGTTGGCAACACGCGCAAGCTAATACAGTATTTAA 456

RESULT 2

US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Clouthier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; TITLE OF INVENTION: OF SALMONELLA
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,788A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043, 403C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEX: 3723836 SEBANDERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-788A-56

Query Match 72.8%; Score 331.8; DB 1; Length 361;
Best Local Similarity 99.4%; Pred. No. 6e-106;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 64 GTCTTCCAAATGAGGCGCGGCGGTAAATCAATACGGGCGGCAATAGTTCCGGCCG 123
Db 1 GTCTTCCAAATGAGGCGCGGCGGTAAATCAATACGGGCGGCAATAGTTCCGGCCG 60
QY 124 GACTCAACGTTGAGCATTTATCAATACGGTCCGCTAAACGCTGCGCTCTGCAAGC 183
Db 61 GACTCAACGTTGAGCATTTATCAATACGGTCCGCTAAACGCTGCGCTCTGCAAGC 120
QY 184 GATGCGCGTAAATCTGAAACGACATTTACCGAGCGGCTTATGTTACGGCGCGATGTA 243
Db 121 GATGCGCGTAAATCTGAAACGACATTTACCGAGCGGCTTATGTTACGGCGCGATGTA 180
QY 244 GGCAGGCGGTGAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 303
Db 181 GGCAGGCGGTGAGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
QY 304 ACCATCGACCAAGTGAAGCTAAATCTCCGATTTATCTGCGGCGCAATACGGCGGT 363
Db 241 ACCATCGACCAAGTGAAGCTAAATCTCCGATTTATCTGCGGCGCAATACGGCGGT 240
QY 364 AACGCGCGGTGTTATCAACCGCATCTGATTC 398
Db 301 AACGCGCGGTGTTATCAACCGCATCTGATTC 335

RESULT 3

US-09-120-927-1/c
; Sequence 1, Application US/09120927
; Patent No. 6262018
; GENERAL INFORMATION:
; APPLICANT: Kim, Jihyun Francis
; APPLICANT: Bear, Steven V.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
; TITLE OF INVENTION: ERWINIA AMYLOVORA AND ITS USE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: P.O. Box 1051, Clinton Square
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/120,927
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/055,108
; FILING DATE: 06-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/1581
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304

QY 178 CAAAGGAGGCCCGCTAAATCTGAAACGACACTTACCGAGCCGTTATGTGAACGGGGC 237
Db 204 CAACTACTGACCGAAGCTAGGTAACTGGAACCTGGGCACTCTGCAGGTTGGGTGCTTC 145
QY 238 GATGTAGGCGCAGGGTGCGGATATATGTACTTATGAACGTACTCGAATGGTTCAGAAAT 297
Db 144 CAGTTCATCATTTTGCATCATCATGTTGGCTGTGCTTTTGACACTTATNGGTCACACACT 85

Db 876 CAGAGTGGTGGACCATTTAGACCAACATGTAGACGACTAGCGCATTAAGAGCCATT 935
Or 340 ACTGTCGCCAATACGGCGGTATATACCC 369
Db 936 GTTGCTCCCAATCGTGGGAGAGCTCC 965

RESULT 10
US-08-916-421B-1
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Built et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
Patent No. 6503729
FILE REFERENCE: P8275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1148948)..(1148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855339)..(855339)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature

```

; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1115881)..(1115881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc feature
; LOCATION: (1664854)..(1664854)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match
Best Local Similarity 6.8%; Score 30.8; DB 4; Length 1664976;
Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
QY 288 TTTCAGAAATATATGCGACCATGACAGCTGGAAGCGTTAAATCCGATTTACTGCGG 347
DB 360315 TTTCATTAATGTTGTCATGATGTCATTAATCAAGAAATTAATCTTTTATTTATTCAT 360374
QY 348 CCAATACCGCGGTAAATAGCGCGGCTGTTAATCAGACCGCATCTGATTCACGCGTAAAT 407
DB 360375 CAACGCTAAAAACAAGAAAGCTTTATGATATGATGATCTTTTAAATCAATTTAAAG 360434
QY 408 GGTGCTGAGGTGGTT 425
DB 360435 GATGCTTAGGTTTAT 360452

```

```

RESULT 11
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1

```

```

; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

```

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 3; Length 4403765;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 45 TGGCAGTCTCTGCTGCGCTGCTTCACAAATGGGCGCGCGGTATCATTAACGGCGG 104
DB 1357218 TTGCCAGCGCTGCTGACGCGCGCGCTCACACGCGCGCTGAGAGATCCCTTGA 1357159
QY 105 CGGCATATGTTCCGCGCCGAGCTCAACGTTGACATTATCAATACGTTCCGCTAACGC 164
DB 1357158 AAACACCGCATGCTCCATACGAGCCACGATTAAGTGTGCGCATGTATCTGTGCGCG 1357099
QY 165 TGGCCTTCTCTG 177
DB 1357098 CCCGACGTGCTG 1357086

```

```

RESULT 12
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

```

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 3; Length 4411529;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
QY 45 TGGCAGTCTCTGCTGCGCTGCTTCACAAATGGGCGCGCGGTATCATTAACGGCGG 104
DB 1357750 TTGCCAGCGCTGCTGACGCGCGCGCTCACACGCGCGCTGAGATCCCTTGA 1357691
QY 105 CGGCATATGTTCCGCGCCGAGCTCAACGTTGACATTATCAATACGTTCCGCTAACGC 164
DB 1357690 AAACACCGCATGCTCCATACGAGCCACGATTAAGTGTGCGCATGTATCTGTGCGCG 1357631
QY 165 TGGCCTTCTCTG 177
DB 1357630 CCCGACGTGCTG 1357618

```

```

RESULT 13
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; PRIOR FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10904
;; LENGTH: 1008
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904

Query Match 6.5%; Score 29.8; DB 4; Length 1008;
Best Local Similarity 54.0%; Pred. No. 3;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGCGCGCGCTATATCAACG 100
Db 141 TGTATGCTGTGCTGCGTACCGAGCGCGCTTGGCGCGCGCGCGCTGG 200
Oy 101 GCGCGCGCAATAGTTCGCGCGCGCGACTCAAGCTTGAGCATTTATCAGTACG 153
Db 201 GCGGGAGAGAGAGCGCGCGCGCGCTGCGCGCGCTTGATCAAGCAGAT 253

RESULT 14
US-09-252-991A-10818/c
;; Sequence 10818, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136
;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10818
;; LENGTH: 1077
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10818

Query Match 6.5%; Score 29.8; DB 4; Length 1077;
Best Local Similarity 54.0%; Pred. No. 3.1;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGCGCGCGCTATATCAACG 100
Db 530 TGTATGCTGTGCTGCGTACCGAGCGCGCTTGGCGCGCGCGCGCTGG 471
Oy 101 GCGCGCGCAATAGTTCGCGCGCGCGACTCAAGCTTGAGCATTTATCAGTACG 153
Db 470 GCGGGAGAGAGAGCGCGCGCGCGCTGCGCGCGCTTGATCAAGCAGAT 418

RESULT 15
US-09-252-991A-10865
;; Sequence 10865, Application US/09252991A
;; Patent No. 6551795
;; GENERAL INFORMATION:
;; APPLICANT: Marc J. Rubenfield et al.
;; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;; FILE REFERENCE: 107196.136

;; CURRENT APPLICATION NUMBER: US/09/252,991A
;; CURRENT FILING DATE: 1999-02-18
;; PRIOR APPLICATION NUMBER: US 60/074,788
;; PRIOR FILING DATE: 1998-02-18
;; PRIOR APPLICATION NUMBER: US 60/094,190
;; PRIOR FILING DATE: 1998-07-27
;; NUMBER OF SEQ ID NOS: 33142
;; SEQ ID NO 10865
;; LENGTH: 1521
;; TYPE: DNA
;; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10865

Query Match 6.5%; Score 29.8; DB 4; Length 1521;
Best Local Similarity 54.0%; Pred. No. 3.7;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 41 TTCTGCGAGTCTCTGCGTGGCGCTTCCCAATAGGCGCGCGCTATATCAACG 100
Db 170 TGTATGCTGTGCTGCGTACCGAGCGCGCTTGGCGCGCGCGCGCTGG 229
Oy 101 GCGCGCGCAATAGTTCGCGCGCGCGACTCAAGCTTGAGCATTTATCAGTACG 153
Db 230 GCGGGAGAGAGAGCGCGCGCGCGCTGCGCGCGCTTGATCAAGCAGAT 282

Search completed: March 16, 2004, 04:36:14
Job time : 70.6647 secs

THIS PAGE BLANK (USPTO)

Result	No.	Score	Query	Match	length	DB	ID	Description
C	1	34.2	7.5	616	12	US-10-42z-559-	88945	Sequence 88945, App1
C	2	34	7.5	1344	9	US-09-833-684-	6	Sequence 6, App1
C	3	34	7.5	1344	9	US-09-880-371-	6	Sequence 6, App1
C	4	34	7.5	1344	9	US-09-879-248-	5	Sequence 5, App1
C	5	34	7.5	1344	14	US-10-010-350-	6	Sequence 6, App1
C	6	34	7.5	1344	15	US-10-441-736-	5	Sequence 5, App1
C	7	34	7.5	2886	9	US-09-801-368-	131	Sequence 131, App
C	8	33.4	7.3	7766	12	US-10-22z-566-	3	Sequence 3, App1
C	9	33.4	7.3	7766	14	US-10-22z-182-	3	Sequence 3, App1
C	10	33.4	7.3	7766	14	US-10-143-024-	3	Sequence 3, App1
C	11	33.2	7.3	1662	14	US-10-15z-751-	483	Sequence 3483, App1
C	12	33.2	7.3	90256	8	US-10-126-611-	1	Sequence 1, App
C	13	32.8	7.2	1059	9	US-09-738-626-	482	Sequence 482, App
C	14	32.8	7.2	1113	9	US-09-738-626-	481	Sequence 481, App
C	15	32.8	7.2	3309400	9	US-09-738-626-	1	Sequence 1, App1

16	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appl1
17	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appl1
18	32.2	7.1	1083	15	US-10-369-493-33133	Sequence 33133, A
19	32.2	7.1	14830	15	US-10-282-1228-27869	Sequence 27869, B
20	31.8	7.0	930	9	US-09-815-242-6028	Sequence 6028, Ap
21	31.8	7.0	930	10	US-10-282-1228-20335	Sequence 20335, A
22	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
23	31.8	7.0	978	15	US-10-369-493-24365	Sequence 24365, A
24	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appl1
25	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appl1
26	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appl1
27	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, Appl1
28	31.2	6.8	2394	12	US-10-282-1228-26837	Sequence 26837, A
29	31	6.8	525	10	US-09-818-995-32144	Sequence 32144, A
30	31	6.8	2751	15	US-10-398-221-3563	Sequence 3563, Ap
31	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
32	30.8	6.8	1575	10	US-09-873-3676-255	Sequence 255, Appl1
33	30.8	6.8	1164	14	US-10-086-510-1	Sequence 1, Appl1
34	30.6	6.7	1164	12	US-10-282-1228-13500	Sequence 13500, A
35	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appl1
36	30.6	6.7	2301	9	US-09-815-242-9844	Sequence 4844, Ap
37	30.6	6.7	2301	9	US-09-815-242-9042	Sequence 9042, Ap
38	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App
39	30.4	6.7	392	12	US-10-282-1228-25231	Sequence 25231, A
40	30.4	6.7	992	12	US-10-425-114-31747	Sequence 31747, A
41	30.2	6.6	25220	11	US-09-984-429-364	Sequence 364, App
42	30.2	6.6	379	12	US-10-424-599-102524	Sequence 102524, A
43	30.2	6.6	654	12	US-10-425-114-28646	Sequence 28646, A
44	30.2	6.6	654	12	US-10-282-1228-24039	Sequence 24039, B
45	30.2	6.6	3222	12	US-10-282-1228-6780	Sequence 6780, Ap

ALIGNMENTS

```

1
US-10-424-599-88945/c
Sequence 88945, Application US/10424599
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
APPLICANT: Cao Yongwei
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 88945
LENGTH: 616
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1) ..(616)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

```

Query Match	7.5%	Score 34.2;	DB 12;	Length 616;
Best Local Similarity	50.9%;	Pred. No. 0.46;		
Matches 81; Conservative	0;	Mismatches 78;	Indels 0;	Gaps 0

Oy 10 GGGCGGCAATAGTTCCGGCCCGAGCTCAAGGTTAGAGATTATCACTACGGTTCGCTA 160
 Db 556 GGGCCCAAGGATTAACCCCTTGATTCACAGCCGGGTCATGTGTGATGATGATGTTG 497
 Oy 161 ACGCTGCGTTCTCTCTGCMAAGCATGCTCCCGTAAATCTGAAGACACATTACCCAGCG 220
 Db 496 ATATATGTATATATAGAGATGATGAGGATGGAGAGTGACCCMAAGATAGTAAACGGG 437

QY 221 GTTATGTACGCGCCGATGTAGCCAGGTCGGATA 259
DB 436 GTAATGTACTAGAGGTATGACAGGAGGTTGGGTA 398

RESULT 2
US-09-835-684-6/c
Sequence 6, Application US/09835684
Patent No. US2002001937A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Qiu, Dewen
APPLICANT: Remick, Dean
TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
FILE REFERENCE: 21829/71
CURRENT APPLICATION NUMBER: US/09/835,684
CURRENT FILING DATE: 2001-04-16
PRIOR APPLICATION NUMBER: 60/198,359
PRIOR FILING DATE: 2000-04-19
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-835-684-6

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTGTGCTGCGCTGCTTCACATGCGGCGCGGTAATCATACGCGCGCAAT 111
DB 209 GCTCCGCTGCGCATTAATGCTGCGCATGACAGTCTTAACAGTTCCGCAAT 150
QY 112 AGTCCGCGCGCGAGCTCAACGTTGAGCATTTATCAGTACGCTCCGTAACGCTGCGCTT 171
DB 149 AATTGAGCCATTGCTCATATGTTGCCGATCGATGAGGTTGTTCCCAACGAGAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 3
US-09-880-371-6/c
Sequence 6, Application US/09880371
Patent No. US20020059658A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Derocher, Jay
TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
FILE REFERENCE: 21829/91
CURRENT APPLICATION NUMBER: US/09/880,371
CURRENT FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: 60/211,585
PRIOR FILING DATE: 2000-06-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-880-371-6

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTGTGCTGCGCTGCTTCACATGCGGCGCGCGGTAATCATACGCGCGCAAT 111
DB 209 GCTCCGCTGCGCATTAATGCTGCGCATGACAGTCTTAACAGTTCCGCAAT 150
QY 112 AGTCCGCGCGCGAGCTCAACGTTGAGCATTTATCAGTACGCTCCGTAACGCTGCGCTT 171
DB 149 AATTGAGCCATTGCTCATATGTTGCCGATCGATGAGGTTGTTCCCAACGAGAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-879-248-5/c
Sequence 5, Application US/09879248
Patent No. US20020062500A1
GENERAL INFORMATION:
APPLICANT: Fan, Hao
APPLICANT: Wei, Zhong-Min
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
FILE REFERENCE: 21829/81
CURRENT APPLICATION NUMBER: US/09/879,248
CURRENT FILING DATE: 2001-06-12
PRIOR APPLICATION NUMBER: 60/212,211
PRIOR FILING DATE: 2000-06-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-879-248-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTGTGCTGCGCTGCTTCACATGCGGCGCGCGGTAATCATACGCGCGCAAT 111
DB 209 GCTCCGCTGCGCATTAATGCTGCGCATGACAGTCTTAACAGTTCCGCAAT 150
QY 112 AGTCCGCGCGCGAGCTCAACGTTGAGCATTTATCAGTACGCTCCGTAACGCTGCGCTT 171
DB 149 AATTGAGCCATTGCTCATATGTTGCCGATCGATGAGGTTGTTCCCAACGAGAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-10-010-390-6/c
Sequence 6, Application US/10010390
Publication No. US20030104979A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Leon, Ernesto
TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
FILE REFERENCE: 21829/111
CURRENT APPLICATION NUMBER: US/10/010,390
CURRENT FILING DATE: 2001-11-05
PRIOR APPLICATION NUMBER: 60/248,169
PRIOR FILING DATE: 2000-11-13
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora

Query Match	7.3%	Score 33.4	DB 12	Length 7766
Best Local Similarity	47.4%	Pred. No. 3.3		
Matches 100	Conservative	0	Mismatches 111	Indels 0
			Gaps 0	
157	GCTACCGCTGCTGCTGCAAGGAGGCCGTAAATCTGAAGACATTACCG	216		

```
Db      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTTCAAAATCCG 4491
Qy      217 AGCGGTTATGTTACGCGCCCGATGTAGCCAGGCGGATTAATGTAATTAAGTGAACG 276
Db      4492 ATTGCGCTGCTAATTGGGAAGGATATGAGATGGTGCGGATTTGGTAAGATGAGATG 4551
Qy      277 ACTCAGATGTTTTCAGAAATATGCAACCATGACCACTGAGCAAGCTTAAACTCCGAT 336
Db      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGAAAG 4611
Qy      337 ATTACTGTGGCCCAATACGCGGTAATACG 367
Db      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642
```

```
RESULT 9
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: JING-REN, STEVEN J.
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD4
; CURRENT APPLICATION NUMBER: US/10/222.162
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-162-3
```

```
Query Match      7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy      157 GCTAACGCTGGCTTGTCTGTCGAAAGCATGCGCCGTAATCTGAAGACCATTAACCCAG 216
Db      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTTCAAAATCCG 4491
Qy      217 AGCGGTTATGTTACGCGCCCGATGTAGCCAGGCGGATTAATGTAATTAAGTGAACG 276
Db      4492 ATTGCGCTGCTAATTGGGAAGGATATGAGATGGTGCGGATTTGGTAAGATGAGATG 4551
Qy      277 ACTCAGATGTTTTCAGAAATATGCAACCATGACCACTGAGCAAGCTTAAACTCCGAT 336
Db      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGAAAG 4611
Qy      337 ATTACTGTGGCCCAATACGCGGTAATACG 367
Db      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642
```

```
RESULT 10
US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, STEVEN J.
; APPLICANT: HARDHAM, JOHN M.
```

```
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: UTSH:234USD1
; CURRENT APPLICATION NUMBER: US/10/143.024
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-143-024-3
```

```
Query Match      7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy      157 GCTAACGCTGGCTTGTCTGTCGAAAGCATGCGCCGTAATCTGAAGACCATTAACCCAG 216
Db      4432 GTTACGCGTGGCGCTGCTGTGAGCAGATGAGAGAACCTGCGAGGCTTCAAAATCCG 4491
Qy      217 AGCGGTTATGTTACGCGCCCGATGTAGCCAGGCGGATTAATGTAATTAAGTGAACG 276
Db      4492 ATTGCGCTGCTAATTGGGAAGGATATGAGATGGTGCGGATTTGGTAAGATGAGATG 4551
Qy      277 ACTCAGATGTTTTCAGAAATATGCAACCATGACCACTGAGCAAGCTTAAACTCCGAT 336
Db      4552 AAGAAAGATGATCAATGTTCTGCTGCTATTGCTTTGAGGGGGATGGCTAAGGATGAAAG 4611
Qy      337 ATTACTGTGGCCCAATACGCGGTAATACG 367
Db      4612 TTTGCTGTGAAGATATGATGATGTGAGAAAG 4642
```

```
RESULT 11
US-10-156-761-3483/C
; Sequence 3483, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156.761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3483
; LENGTH: 1662
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1662)
US-10-156-761-3483
```

```
Query Match      7.3%; Score 33.2; DB 14; Length 1662;
```

Best Local Similarity 52.1%; Pred. No. 1.7;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTGCGGCGGCTGTTCCACAAATGGGCGGCGGTAATCATTAACGGCGCGCA 109
Db 590 GCGGCGCGGCTGAGCATGATCAGATCGCGCGTTCATCGTCAACGGCGAGCGCG 531
Qy 110 ATAGTCCGGCGGCGGACTCAACGTTGAGCATTATCACTAGCGTTCCGCTAACGTCGCG 169
Db 530 TGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 471
Qy 170 TTGCTTGCAAGCGATGCGCG 191
Db 470 TTGAACCTTCTTGCGCGGCGACGG 449

RESULT 12
US-10-156-761-1

Sequence 1, Application US/10156761
Publication No. US20030119018A1

GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 7.3%; Score 33.2; DB 14; Length 9025608;
Best Local Similarity 52.1%; Pred. No. 1.3e+02;
Matches 74; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 GTGCTGCGGCGGCTGTTCCACAAATGGGCGGCGGTAATCATTAACGGCGCGCA 109
Db 4331781 GCGGCGGCGGCTGAGCATGATCAGATCGCGCGGCTTCATCGTCAACGGCGAGCGCG 4331840
Qy 110 ATAGTCCGGCGGCGGACTCAACGTTGAGCATTATCACTAGCGTTCCGCTAACGTCGCG 169
Db 4331841 TGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4331900
Qy 170 TTGCTTGCAAGCGATGCGCG 191
Db 4331901 TTGAACCTTCTTGCGCGGCGACGG 4331922

RESULT 13
US-09-738-626-482

Sequence 482, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO

APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 482
LENGTH: 1059
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

Query Match 7.2%; Score 32.8; DB 9; Length 1059;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATGCCCGTAATGTGAAGACCATTAACCGGCGGTTATGTTAAGCGGCGGATGTA 243
Db 76 GAGGACGTTTCCACCGGAGAGACACGACTGCAAGCTCTTCGCTTAAGCATCGATGCA 135
Qy 244 GCGGCGGCGGATTAATGACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 303
Db 136 GCGCGTAAAGAAAGTAAACATCACCGTCTACACCTGAGCGTGAAGAAAGTGAAGTGAAG 195
Qy 304 ACCATGACCGAGTGAAGCGGTAAGAACTCCGATTTACTGCGGCGCAATACGGCGGTAT 363
Db 196 ATCAACAAGGCGTTCATGAGAAACCAACCCAGATATTAAGTTGAGGTGATCGCGCGTGGT 255
Qy 364 AACGCGCGGCTGTTAATCAGACCGCATCTGA 395
Db 256 ACTGGCATCTGACTGCTCGCATTTGAAGCTGA 287

RESULT 14
US-09-738-626-481/c

Sequence 481, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 481

LENGTH: 1113
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

Query Match 7.2%; Score 32.8; DB 9; Length 1113;
Best Local Similarity 47.2%; Pred. No. 1.9; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 184 GATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCGCATGTA 243
DB 972 GAGGACGTTTCACCGCAGAGCCACACGACGCTCTCCGCTAACGATCCGATGCA 913
QY 244 GCGCAGGTTGGGATTAATAGTACTTGAAGTCAAGTGGTTTCAAGAAATATGCC 303
DB 912 GCCGCTGAAAAAGTAAACATCACTGCTTACCTTGTAGCCTGAGGAAAAAGTGCATGAG 853
QY 304 ACCATGACGAGTGAGACGCTAAAACTCCGATATTACTGTGGGCAATACGGCGTAAAT 363
DB 852 ATCAACAAAGGGGTTGATGGAAGCCACCAATATGAGTTGAGTGTACCGCGCTGCT 793
QY 364 AACGCGCGCTGTTTAACTCAACGCGCATCTGA 395
DB 792 ACTGCGCATCTGACTGCTGCGCATTGAACTGA 761

RESULT 15

US-09-738-626-1/c
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738, 626
CURRENT FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 7.2%; Score 32.8; DB 9; Length 3309400;
Best Local Similarity 47.2%; Pred. No. 1.2e+02;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 184 GATGCCCGTAATCTGAAACGACATTAACGAGCGGTTATGTAACGGCGCATGTA 243
DB 451820 GAGGACGTTTCACCGCAGAGCCACACGACGCTCTCCGCTAACGATCCGATGCA 451761
QY 244 GCGCAGGTTGGGATTAATAGTACTTGAAGTCAAGTGGTTTCAAGAAATATGCC 303
DB 451760 GCCGCTGAAAAAGTAAACATCACTGCTTACCTTGTAGCCTGAGGAAAAAGTGCATGAG 451701
QY 304 ACCATGACGAGTGAGACGCTAAAACTCCGATATTACTGTGGGCAATACGGCGTAAAT 363

DB 451700 ATCAACAGGCGTTCAATGGAAGCCAAACCAGATATTGAGGTTGAGGTGATCCGCGCTGCT 451641
QY 364 AACGCGCGCTGTTTAACTGAGACCGCATCTGA 395
DB 451640 ACTGCGATCTGACTGCTGCGCATTGAACTGA 451609

Search completed: March 17, 2004, 08:14:59
Job time: 419.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-1
Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
EST:
1: em_estha:*
2: em_esthm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hcc:*
9: gb_estcl:*
10: gb_estc2:*
11: gb_hcc:*
12: gb_estc3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	255	55.9	680	12	BU618688
C 2	37.4	8.2	484	9	AV430994
C 3	35.2	7.7	813	28	B2204853
C 4	34.6	7.6	702	12	BG441624

5	34.6	7.6	937	12	BG444255	BG444255 GA_Ea002
6	34.2	7.5	296	9	AV058630	AV058630 AV058630
7	34.2	7.5	523	12	BU334624	BU334624 BU334624
8	34.2	7.5	1036	12	BG671547	BG671547 DRNBUB05
C 9	34	7.5	500	14	CA711477	BB581987 BB581987
C 10	34	7.5	500	14	CA711477	CA711477 wdk2c_pk0
C 11	34	7.5	811	28	B2469000	B2469000 BOO0058TF
C 12	33.8	7.4	365	9	AA746477	AA746477 nmw2e02_8
C 13	33.8	7.4	358	14	CD374421	CD374421 TMBMFC2H
C 14	33.8	7.4	559	13	BU765156	BU765156 sa623f11.
C 15	33.8	7.4	941	29	CG178280	CG178280 PUUDF79TD
C 16	33.8	7.4	1098	12	CNS02GQP	AL196666 Tetracodon
C 17	33.6	7.4	594	12	BU335653	BU335653 BU335653
C 18	33.6	7.4	653	12	BU336964	BU336964 BU336964
C 19	33.6	7.4	712	12	CF437982	CF437982 EST674327
C 20	33.4	7.3	505	6	AL809250	AL809250 Tt1t1cum
C 21	33.4	7.3	679	28	CC961526	CC961526 BOIGS10TF
C 22	33.4	7.3	695	28	BH663951	BH663951 BOMBL85TR
C 23	33.4	7.3	699	12	B1959328	B1959328 HVSME001
C 24	33.4	7.3	700	28	BH651371	BH651371 BOHXV71TF
C 25	33.4	7.3	709	28	BH690914	BH690914 BOHXK64TF
C 26	33.4	7.3	715	13	BU444683	BU444683 603765715
C 27	33.4	7.3	785	28	BH578793	BH578793 BOGLX20TR
C 28	33.4	7.3	815	28	BH461716	BH461716 BOHIV03TR
C 29	33.4	7.3	827	28	BH438448	BH438448 BOGV028TR
C 30	33.4	7.3	830	28	BH715216	BH715216 BOMIM42TR
C 31	33.4	7.3	835	28	BH575194	BH575194 BOH0277TF
C 32	33.4	7.3	853	28	BH578402	BH578402 BOHNV61TF
C 33	33.4	7.3	1026	28	BZ458784	BZ458784 BOMBL14TF
C 34	33.2	7.3	589	28	BZ300068	BZ300068 KD0510.p1
C 35	33.2	7.3	625	10	BF501499	BF501499 AT16857.5
C 36	33.2	7.3	629	9	AU295325	AU295325 AU295325
C 37	33.2	7.3	788	9	AU139321	AU139321 AU139321
C 38	33.2	7.3	1646	10	BF579120	BF579120 602096054
C 39	33	7.2	456	12	BU331339	BU331339 BU331339
C 40	33	7.2	524	12	BU367372	BU367372 BU367372
C 41	33	7.2	547	12	BU333958	BU333958 BU333958
C 42	33	7.2	553	12	BU336903	BU336903 BU336903
C 43	33	7.2	569	12	BU334047	BU334047 BU334047
C 44	33	7.2	574	12	BU330328	BU330328 BU330328
C 45	33	7.2	607	12	BU365766	BU365766 BU365766

ALIGNMENTS

RESULT 1
LOCUS BU618688/680 bp mRNA linear EST 01-OCT-2003
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186D22 5', mRNA sequence.
BU618688
BU618688.1 GI:37256713

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Xenopus laevis (African clawed frog)

REFERENCE
AUTHORS
Kohara, Y., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@genes.nig.ac.jp
The information of this clone is available through the following URL.

COMMENT

TITLE
JOURNAL
COMMENT

FEATURES
source http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X116b22"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/note="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN
Query Match 55.9%; Score 255; DB 12; Length 680;
Best Local Similarity 72.4%; Pred. No. 3.8e-62;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 1 ATGAAACCTTTAAAGTCGACGATTCGAGCAATCGTAGTTCTGGCAGTGTCTGCT 60
DB 593 ATGAAACCTTTAAAGTCGACGATTCGAGCAATCGTAGTTCTGGCAGTGTCTGCT 534
QY 61 GGGTCGTCCTCAATGGGCGCGCGCGGTATCTAAACGGCGCGCAATAGTTCCGCG 120
DB 533 GGTGTGTCTCTCAGACGGCGCGCGGTATCTAAACGGCGCGCAATAGTTCCGCG 474
QY 121 CCGGACTCAACGCTTATGATAGTACGAGTTCGCTTACCGCTTACCGCTTACCGCT 180
DB 473 CCAAAATCTGAGCTGAACTTTACAGTACGCTGCGGTATCTGCTTCTGCTGAA 414
QY 181 AGCGATGCGCGCTTAACTTGAAACGCAATTCACCAAGCGGTATGTTACCGCGCGCAT 240
DB 413 ACTGATGCGCGCTTAACTTGAAACGCAATTCACCAAGCGGTATGTTACCGCGCGCAT 354
QY 241 GTAGCGCAGGCTGCGATATAGTATCTATGACGCTGAGTGTTCAGAAATAT 300
DB 353 GTTGTGACGGCTGAGTACGATGCTCAATCTGACCAACGCTGCTTCCGTTAACAGC 294
QY 301 GCCACATCGACCGATGGAACGCTTAAACTCCGATATTACTGCGCCAAATAGCGCGGT 360
DB 293 GCTACTCTTGTATGATGAGTGAACGCGCAAAATCTGAAATGACGTTAAACGTTCCGTTGT 234
QY 361 AATAAGCGCGCGCTGTTAATCAGACCGCATCTGATTCACGCGTAAATGCTGCTGAGTT 420
DB 233 GCGAAGCGTGTGAGTGTGACCAATCTGATCTTCTCCGTCACGACTCAGGTT 174
QY 421 GGTTTGGCAACAGCGCAGCGCTTAAACGAGTATTTAA 456
DB 173 GGTCTTGTGTAACAGCGCAGCGCTCATCAGTACTTAA 138

RESULT 2
AV430994 484 bp mRNA linear EST 23-AUG-2000
LOCUS AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
DEFINITION Pl027e10 r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
1 (bases 1 to 484)
Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Tabata, S.
Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
JOURNAL DNA Res. 7, 223-227 (2000)
MEDLINE 20363100
PUBMED 10907854
COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

FEATURES
source Yana 1532-3; Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
Location/Qualifiers
1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="Pl027e10_r"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN
Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 9.5;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
QY 321 CGCTAAACCTCGATATTACTGTGCGCCATACGCGGTATTAACGCGCGCTGTTAA 380
DB 441 CCCAAACACACAGTGTCCGCGCGCTTCAAGACGCGCTTCAATCGCTGCTGCGCAG 382
QY 381 TCAGACCGCATCTGATTCAGCGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 439
DB 381 GCCAAAGCTGTGATCCAAAGTACTGCTGCGAGGCTGCGGAAGGCGCTGCGCGCA 323

RESULT 3
BZ204853 813 bp DNA linear GSS 11-OCT-2002
LOCUS BZ204853
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-390D21, genomic survey sequence.
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 813)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,
Shyartbeyn, A., Gebregorgis, E., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P., and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 Mb01 segment
Unpublished (1999)
Other GSSs: CH230-390D21.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or erting_information.htm). BAC end
page: http://www.tigr.org/tldb/bac_ends/rat/bac_end_intro.html
Seq primer: T7
Class: BAC ends.

FEATURES
source Location/Qualifiers
1. .813
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHsd/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"

/note="Vector: pTRABAC1.3; Site 1: Mbol; Site 2: Mbol;
 CHORI-230 Rat (BN/Sevhad/MCM) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 813;
 Best Local Similarity 49.0%; Pred. No. 51;
 Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

Qy 178 CAAGCCGATCCCGTAAATCTGAAACGACCATTCACAGAGCGGTTATGTAACGGCGCC 237
 Db 548 CAAGGACCTCCACATCAACACAGACACACTCAAACTAATAGAGAAAAGTACGGAAGCA 607
 Qy 238 GATGAGCCGAGGTGGCGATTAATAGTACTATTAAGCTGACTCAGAAATGTTTCAGAAT 297
 Db 608 TCTGGAACATGGGCACTGGAATAAAATCTCTGAAACAAACCAATGGCTTATGCTCT 667
 Qy 298 AATGCCACCATCGACGATGGAACGCTAAAACTCCGATATTACTGTGCGCAATACGGC 357
 Db 668 AAGATCAAGATCGACCAATGGGATCTCATTAATCTGCAAGCTTCTGTAGCGAAAGACAC 727
 Qy 358 GGTAAATACGCC 369
 Db 728 TGTGTTAGGAC 739

RESULT 4
 BG441624 702 bp mRNA linear EST 15-MAR-2001
 LOCUS GA_Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION
 accession BG441624 GI:13351276
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 702)
 Wing, R.A., Friesch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)

JOURNAL
 COMMENT
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 696.
 Location/Qualifiers
 1..702
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultiivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0014A14f"
 /issue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES

ORIGIN
 Query Match 7.6%; Score 34.6; DB 12; Length 702;
 Best Local Similarity 46.5%; Pred. No. 71;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 181 AGGATGCCCGTAAATCTGAAACGACCATTCACAGAGCGGTTATGTAACGGCGCAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGGCAGCAACATCAAGACACACCAAT 244
 Qy 241 GTAGCCAGGATGCGGATTAATAGTACTATTGAATGACTGACCTCAGAAATGTTTCAGAAT 300
 Db 245 GACAATGTAATGCAATGGAATGAGATGTTACTCAAAAGAAAAACATTCAGATTCAAAT 304
 Qy 301 GCCACCATCGACGATGGAACGCTAAAACTCCGATATTACTGTGCGCAATACGGCGGT 360
 Db 305 GAAATGCTGACAGACCAACACCAACACATGAAAAACATTAATGAAAAATGCGGT 364
 Qy 361 AATAGCCCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATAGTGCCTCAGGT 420
 Db 365 CAGACGGGAAGAAATGCGACTCAGATGACATGACATGAAATGTTGGTCAAGAT 424
 Qy 421 G 421
 Db 425 G 425

RESULT 5
 BG444255 937 bp mRNA linear EST 15-MAR-2001
 LOCUS GA_Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 DEFINITION
 accession BG444255 GI:13353907
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 937)
 Wing, R.A., Friesch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Henry, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)

JOURNAL
 COMMENT
 Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATGAGG
 High quality sequence stop: 804.
 Location/Qualifiers
 1..937
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultiivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0023M14f"
 /issue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_lib="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 937;
 Best Local Similarity 46.5%; Pred. No. 81;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

Qy 181 AGGATGCCCGTAAATCTGAAACGACCATTCACAGAGCGGTTATGTAACGGCGCAT 240
 Db 185 AACATGACGGGAATGCGAATGGGAATGTTGGCAGCAACATCAAGACACACCAAT 244
 Qy 241 GTAGCCAGGATGCGGATTAATAGTACTATTGAATGACTGACCTCAGAAATGTTTCAGAAT 300

Db 245 GACATGTAATGCAATGATGTTACTCAAAAACACAGATTCAGTTCAAT 304
 Qy 301 GCCACCATGACCAAGTGGAAAGCTTAATAAATCCGATATATCTGTGGCCATACGGCGGT 360
 Db 305 GAAATGCTGCAAGAGCAACACCAACATGAAACAAATGAAGAAAATGCGGT 364
 Qy 361 AATAAGCCGCGCTGGTTATCAGACCGATCTGATTCAGGGTAAATGTCAGATT 420
 Db 365 CAGACGGGAGGAATCCACTCAGAGTAACTGACCAACATGAAATGTGTCAGAT 424
 Qy 421 G 421
 Db 425 G 425

RESULT 6
 AV058630 296 bp mRNA linear EST 23-JUN-1999
 LOCUS AV058630
 DEFINITION clone 1810054H02, mRNA sequence.
 ACCESSION AV058630
 VERSION AV058630.1 GI:5158377
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 296)

TITLE
 JOURNAL
 COMMENT
 AUTHORS Akahira, S., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Nishitama, H., Oda, H., Sugahara, Y., Suzuki, H., Shigemoto, Y., Shiraki, T., Sogabe, Y., Tomioka, N., Yagame, S., Yagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
 RIKEN Mouse ESTs
 Unpublished (1999)
 Contact: Chie Owa
 Genome Science Laboratory
 RIKEN

3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
 Tel: 81-298-36-9145
 Fax: 81-298-36-9098
 Email: genome-res@rtc.riken.go.jp
 Thermotabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
 Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES
 source
 1. 296
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="1810054H02"
 /sex="male"
 /tissue_type="pancreas"
 /dev_stage="adult"
 /clone_1ib="Mus musculus pancreas C57BL/6J adult"

ORIGIN
 Query Match 7.5%; Score 34.2; DB 9; Length 296;
 Best local Similarity 53.3%; Pred. No. 63;
 Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 251 GTGGCGATATAGTACTATTGAACTGCAAGTGGTTTCAAGAAATGATCCACCATCG 310
 Db 13 GTGGAGACCAAGTTTCTTAGCTACCAAGGCTGGTAAAGAAAGAGCATCAACTG 72

Qy 311 ACCAGTGAAGCGCTTAAATCCGATATTTACTGTGGCCCAATACGGCGTAAATACGCCG 370
 Db 73 TCCAGTGAAGCAAAATGTAAGAAAGAAATGATTATCCCAAGTGGATTTAGTG 132
 Qy 371 CGCTGTTAATACGA 385
 Db 133 AGATGCCAAATAGA 147

RESULT 7
 B3334624 523 bp mRNA linear EST 05-MAR-2002
 LOCUS B3334624
 DEFINITION B3334624 Dictyostelium discoideum cDNA library, Af dictyostelium
 ACCESSION B3334624
 VERSION B3334624.1 GI:19164754
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum

REFERENCE
 AUTHORS Uzunihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadao Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.

FEATURES
 source
 1. 523
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda47006"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 /clone_1ib="Dictyostelium discoideum cDNA library, Af"

ORIGIN
 Query Match 7.5%; Score 34.2; DB 12; Length 523;
 Best local Similarity 57.1%; Pred. No. 82;
 Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 238 GATGTAGCGCAGGCTGCGATATATGTTACTGTAAGTCAAGATGTTTCAGAAAT 297
 Db 398 GATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
 Qy 298 AATGCCACCATGACCATGATGATGATGATGATGATGATGATGATGATGATGAT 342
 Db 458 AATAACCAACACCAACCAATATATATATATATATATATATATATATATATATAT 502

RESULT 8
 BG671547 1036 bp mRNA linear EST 30-APR-2001
 LOCUS BG671547
 DEFINITION DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
 ACCESSION BG671547
 VERSION BG671547.1 GI:13893646
 KEYWORDS EST.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus

REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus. 1 (bases 1 to 1036)
 Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C.,

[illegible]

Db 200 TAGTTTTTGGTAGGTCGGTTATGC

TITLE	JOURNAL	COMMENT
<p>Kojima, Y., Komio, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Nomazaki, R., Okazaki, T., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takehashi, F., Tanaka, T., Toyota, T., Matsubaki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A., Muramatsu, M. and Hayashizaki, Y.</p> <p>RIKEN Mouse ESTs (Aizawa, K. et al. 2000)</p> <p>Unpublished (2000)</p> <p>Contact: Yoshihide Hayashizaki</p> <p>Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute</p> <p>The Institute of Physical and Chemical Research (RIKEN)</p> <p>1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan</p> <p>Tel.: 81-45-503-9222</p> <p>Fax: 81-45-503-9216</p> <p>Email: genome-res@gsc.riken.go.jp,</p> <p>url: http://genome.gsc.riken.go.jp/</p> <p>Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagao, S., Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.</p> <p>Thermosensitization and thermooxidation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)</p> <p>Itoh, M., Katsunai, T., Akiyama, Y., Shibata, K., Itawa, M., Kawai, J., Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.</p> <p>Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)</p> <p>Carninci, P. and Hayashizaki, Y.</p> <p>High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)</p> <p>Please visit our web site (http://genome.irc.riken.go.jp) for further details.</p> <p>Location/Qualifiers</p> <p>1. 234</p> <p>/organism="Mus musculus"</p> <p>/mol_type="mRNA"</p> <p>/strain="C57BL/6J"</p> <p>/db_xref="taxon:10090"</p> <p>/clone="9030005005"</p> <p>/sex="male"</p> <p>/tissue_type="colon"</p> <p>/dev_stage="adult"</p> <p>/lab_host="DH10B"</p> <p>/clone_id="RIKEN full-length enriched, adult male colon"</p> <p>/notes="Site 1: Sali; Site 2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'</p> <p>GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTNN 3'] cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 185.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGAGAGAGATTCCTGAGTTAATTAATTAATCCCCCCCCCC 3'] cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: Sali; 3' end: BamHI"</p>		

Qy 95 ATACGGGCGGCGCAATGATTCCGCCGGAAGTCAAGTTGACATTTATCAGTACGCTT 154
 Db 140 CGAACCGCGCGGTCTTATGACCTTTGTTCAAAATGTAGGACATTTACCAACCTTT 81
 Qy 155 CCGCTAACGCTGGCTTGTCTGTGCAAGCATGC 188
 Db 80 CCGTTACCAAGCCCGTCCGCCATGTCGGCGG 47

RESULT 10
 CA711477/c 500 bp mRNA linear EST 26-NOV-2002
 LOCUS wdK2c.pk014.03 wdK2c Triticum aestivum cDNA clone wdK2c.pk014.03 5'
 DEFINITION end, mRNA sequence.
 ACCESSION CA711477
 VERSION CA711477.1 GI:25433270
 KEYWORDS EST.
 SOURCE Triticum aestivum (bread wheat)
 ORGANISM Triticum aestivum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 500)
 Tingley, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,
 Miao, G., Caraher, N. and Hanaley, M.K.
 Duront Wheat cDNA Sequence
 Unpublished (2002)
 CONTACT: Scott V. Tingley
 CROP GENETICS
 E. I. Dupont de Nemours and Company
 1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
 TEL: 302-631-2602
 FAX: 302-631-2607
 EMAIL: Scott.V.Tingley@usa.dupont.com
 Seq primer: M13.
 Location/Qualifiers
 1..500
 /organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone="wdK2c.pk014.03"
 /issue_type="kernel"
 /clone_lib="wdK2c"
 /note="Vector: pBluescript SK+, Site 1: EcoRI, Site 2:
 XhoI, Wheat (Triticum aestivum L.) developing kernel, 7
 days after anthesis."

FEATURES
 source
 Query Match 7.5%; Score 34; DB 14; Length 500;
 Best Local Similarity 54.5%; Pred. No. 91;
 Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;
 Qy 15 AGTGGCAGCATTCGACGAATCTAGTTTCTGGAGAGTCTGTGGCTGCTGTCACACA 74
 Db 274 AGNTGNTGACAGTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 215
 Qy 75 ATGGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 126
 Db 214 TGTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 163

RESULT 11
 BZ469000/c 811 bp DNA linear GSS 13-DEC-2002
 LOCUS BOOA0587F BO_1.6_2_KB tot Brassica oleracea genomic clone BOOA058,
 DEFINITION genomic survey sequence.
 ACCESSION BZ469000
 VERSION BZ469000.1 GI:26764546
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

REFERENCE rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 1 (bases 1 to 811)
 Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 Whole genome shotgun sequencing of Brassica oleracea
 Unpublished (2001)
 JOURNAL Other GSSs: BOOA0587F
 COMMENT Contact: Chris Town

TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.

FEATURES
 source
 Location/Qualifiers
 1..811
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="T01000DH3"
 /db_xref="taxon:3712"
 /clone="BOOA058"
 /clone_lib="BO_1.6_2_KB tot"
 /note="Vector: pHOSt1; Site 1: BstXI, 1.6-2 kb sheared
 total DNA inserted into pHOSt1 using BstXI linkers"

ORIGIN
 Query Match 7.5%; Score 34; DB 28; Length 811;
 Best Local Similarity 48.9%; Pred. No. 1.1e+02;
 Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 Qy 176 TSCAAAGCGATGCCCGTAAATCTGAAAGACCATTAACAGAGCGGTATGACGGC 215
 Db 610 TGTAGGCAACATTTGAAACCTGCTTCCATTTATCTACTGTTATGACCAATGCT 551
 Qy 236 CCGATGTAGGCGCAGGCGCGGATTAATGACTATGAACTGACTCAGAAATGTTGACGA 295
 Db 550 TTGATCTAATCACTGATGATGTTGAGCTGCTCTGCTGTTAGAGATGTTTAAAT 491
 Qy 296 ATAAATGCCACATGACCACTGGAACGCTTAAACCTCCATTTACTGTGCGCCAAATAG 355
 Db 490 ACTTGTCACTTATTAATGAATAATGAAATACCAATACCTGATTAACACTTATTAACCA 431
 Qy 356 GCGGTA 361
 Db 430 AAGATA 425

RESULT 12
 AA746477/c 365 bp mRNA linear EST 27-JAN-1998
 LOCUS mw62602.s1 NCI_CGAP CCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
 DEFINITION similar to TR:000555 000555 NEURONAL CALCIUM CHANNEL ALPHA 1A
 SUBUNIT ISOFORM A-1; contains element TAR1 repetitive element ;,
 mRNA sequence.
 ACCESSION AA746477
 VERSION AA746477.1 GI:2786463
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 365)
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456

Sequence: 1 atgaaacaacatcgtatc.....ctgcgtcaccacgctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_8ts:*
12: gb_8v:*
13: gb_un:*
14: gb_vr:*
15: gb_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_8ts:*
27: em_un:*
28: em_vl:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_other:*
41: em_hcg_hum:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	2067	1	SEU43280
2	452.8	99.3	5103	1	STAJ2301
3	452.8	99.3	22411	1	AB008749
4	451.2	98.9	254050	1	AL627269
5	451.2	98.9	301983	1	AB016840
6	382.8	83.9	1048	1	STACRBA
7	348.4	76.5	2889	1	CPSP15700
8	334.4	73.3	10370	1	AB015131
9	332.8	73.0	4680	1	ECCSGABDG
10	332.8	73.0	10346	1	AE000205
11	332.8	73.0	15047	1	D90741
12	332.8	73.0	282504	1	AB016981
13	332.8	73.0	306358	1	AB016759
14	331.2	72.6	1711	1	AF275733
15	331.2	72.6	10190	1	AE005315
16	331.2	72.6	327773	1	AP002554
17	313.4	68.7	2920	1	CFR515701
18	271	59.4	2883	1	ESAS15702
19	184	40.4	1212	1	EC0131756
20	47.8	10.5	10709	1	AE015532
21	47.4	10.4	301214	1	AE016786
22	44.2	9.7	11160	1	AE008209
23	44.2	9.1	11476	1	AE009405
24	41.4	9.1	42602	2	AC116971
25	41.1	9.0	7218	6	16494
26	40.6	8.9	62381	5	BX640594
27	40.4	8.9	168457	2	AC111425
28	40.4	8.9	178934	2	AC127966
29	40.2	8.8	193553	10	AC124196
30	39.8	8.7	295500	1	AP005954
31	38.2	8.4	298406	3	CE17588A
32	37.8	8.3	174732	2	AL136161
33	37.8	8.3	191567	9	AL138702
34	37.6	8.2	2000	6	AX655393
35	37.4	8.2	120937	10	AL591911
36	37.4	8.2	225016	2	AC063967
37	37.2	8.2	42310	3	CBRG0185
38	37	8.1	164504	2	AC123240
39	36.8	8.1	252420	3	AB014841
40	36.6	8.0	92378	5	BX537249
41	36.6	8.0	110000	2	AC095863_05
42	36.6	8.0	196389	2	AC113159
43	36.6	8.0	230058	2	AC108229
44	36.6	8.0	254497	2	AC119520
45	36.6	8.0	323930	1	AP003194

ALIGNMENTS

RESULT 1
LOCUS SEU43280
DEFINITION Salmonella enteritidis agfBAC operon: fimbrin-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and AgfC (agfC) genes, complete cds.
ACCESSION U43280
VERSION U43280.1
KEYWORDS GI:1184712
SOURCE Salmonella enteritidis
ORGANISM Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE 1 (bases 1 to 2067)
Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, E.C.,

Pred. No. is the number of results predicted by chance to have a

TITLE Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I. and Kay, W.W.
DNA-based diagnostic tests for *Salmonella* species targeting *agfa*,
the structural gene for thin, aggregative fimbriae
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)

JOURNAL MEDLINE 94013373
PUBMED 8104955

REFERENCE 2 (bases 1 to 2067)
AUTHORS Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and Kay, W.W.
Salmonella enteritidis agfBAC operon encoding thin, aggregative fimbriae
J. Bacteriol. 178 (3), 662-667 (1996)

JOURNAL MEDLINE 96146512
PUBMED 8550497

REFERENCE 3 (bases 1 to 2067)
AUTHORS Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
Direct Submission
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg., Victoria, BC V8W 3P6, Canada

FEATURES
source Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
/gene="agfB"
696..1151
/note="fimbria-like protein precursor"
/codon_start=1
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MNKLLFMMLTLLGPGIATNTYDGLARSEYFAVELSKSPFN QAAIQVGTNSNARVREGSKLSTVSGEGNNRPAKVDGNTFYIETGNANDA SISQAYGNSAIAIKQSGNKANITQYGTQTAIVVQKSHMALRYTOR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids; fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=1
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLTKYAAFAIIVSGSLAGVDPWGCGGNNHGGNSGSDST IDQWNAKNSDITTYGQYGGNNALVNOTASDSVVRQVGFNNATANYR"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"

sig_peptide
mat_peptide
stem_loop
gene

CDS
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=1
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTLLLAALNSQITFTTQGGDIYTVIROYLNEPCYQVQIL SVRQVGGQSHYDQKTLSPANQPIELSRISVSISSDSKVIITVSDGSLHSQ MPESAQ"

ORIGIN

Query Match	Best Local Similarity	Score	DB 1;	Length	2067;
Matches	456;	Conservative	0;	Mismatches	0;
				Indels	0;
				Gaps	0;
1	ATGAAAACAAATTTGTTATTTATGATGTGACAAATCTGGTGGCCCTGGGATTGCAAC	60			
696	ATGAAAACAAATTTGTTATTTATGATGTGACAAATCTGGTGGCCCTGGGATTGCAAC	755			
61	GCACAAATTTATGATCTGCTCGTTCAGAAATTAATTTTGGGTAATGAATTAAGCAAG	120			
756	GCACAAATTTATGATCTGCTCGTTCAGAAATTAATTTTGGGTAATGAATTAAGCAAG	815			
121	TCCTCATTTAATCAGCGCCCATTTATTTGTCAGTCCGACGATATATGTGCGACAGTA	180			
816	TCCTCATTTAATCAGCGCCCATTTATTTGTCAGTCCGACGATATATGTGCGACAGTA	875			
181	CGCCGAGGAGTCAAACTATTGTCCGTAATTTCCAGAGAGAGAAATTAATCGGCG	240			
876	CGCCGAGGAGTCAAACTATTGTCCGTAATTTCCAGAGAGAGAAATTAATCGGCG	935			
241	AAAGTCACGAGGAGGAAATTAATTTGCTGATATTTAGCAAAACGGCAATGCCAAC	300			
936	AAAGTCACGAGGAGGAAATTAATTTGCTGATATTTAGCAAAACGGCAATGCCAAC	995			
301	GATGCCGATATTCGCAAAACGGCTTACGTAATTAATGACGATATTAATCAGAAAGTTCT	360			
996	GATGCCGATATTCGCAAAACGGCTTACGTAATTAATGACGATATTAATCAGAAAGTTCT	1055			
361	GAAATTAAGGCAATTTTCCAGTAACGTAACGTAACGTAACGTAACGTAACGTAAC	420			
1056	GAAATTAAGGCAATTTTCCAGTAACGTAACGTAACGTAACGTAACGTAACGTAAC	1115			
421	CAGTCCGATATGCTATTTGCGTCAACCAACGCTTA 456				
1116	CAGTCCGATATGCTATTTGCGTCAACCAACGCTTA 1151				

RESULT 2
STAJ2301 5103 bp DNA linear BCT 15-NOV-2000
LOCUS *Salmonella typhimurium* cs9G, cs9F, cs9E, cs9D, cs9B, cs9A, and k
DEFINITION cs9G genes.
ACCESSION AJ002301
VERSION AJ002301.1 GI:2739232
KEYWORDS cs9A gene; cs9B gene; cs9C gene; cs9D gene; cs9E gene; cs9F gene;
SOURCE *Salmonella typhimurium*
ORGANISM *Salmonella typhimurium*
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.
REFERENCE 1 Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.
TITLE Curli fibers are highly conserved between *Salmonella typhimurium*
and *Escherichia coli* with respect to operon structure and regulation
JOURNAL MEDLINE 98117058
PUBMED 9457880
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling, U.

TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
Source
1. .5103
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
CDS
<1..322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translation="CTLSGMVMAVLIIHGSALAPHLIEIVSLGTIVAFIMCIQAKO
LLISFVPCTFIGACATFAGQDMRLVLPSLALGLIFGYAMNCSGLMASRRRHSANT
AVTK"
gene
complement(361..1194)
/gene="csbg"
CDS
complement(361..1194)
/gene="csbg"
/function="lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Csgb protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAVLISGCTTAPPKQAKFTLMPRAQSYKDLTHLPA
NLGKIPVSVYIDETGQFKPYSPASNSTATMLVTALKDSHFIFLEQGHQ
PLNBERKIRAAQNGTVAAMNRIPLQSLTANIMVSGSIIGSNVSGGVARYG
IGADTQYLDQIIVNLRVNVSTGEIILSVSTKTLISYEQAGVPRPIDQRLBEE
IGTNSRPVMLCLMSALETGYIFLINDIGDLNDLQNKADRONDLVKRELVSPE
S"
gene
complement(1221..1637)
/gene="csgf"
CDS
complement(1221..1637)
/gene="csgf"
/codon_start=1
/transl_table=1
/product="Csgf protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVKAAYVLLMRSPLTWAGNMTFOFNPNGPNNNGSLFNLAS
AQONSYKDPAYNDGIELTSPALDNFTQALQSQILGGLTNINTGKRGKRVNTDPII
DIARDQQLQNLVDRKTRSTIEVSGLTQSTDF"
gene
complement(1664..2059)
/gene="csge"
CDS
complement(1664..2059)
/gene="csge"
/codon_start=1
/transl_table=1
/product="Csg protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWIVAELLFATGNLHANEVEVEFGLLTDHTVSSIGHB
FYRAFSDKWESEYGLNLTINERPARGSGWTTITVNDVIFQFLFPKRDPEKTVVF
ALAQTEBALNRQIDQLTSLSDLRDF"
gene
complement(2064..2714)
/gene="csgd"
CDS
complement(2064..2714)
/gene="csgd"
/function="transcriptional regulator"
/codon_start=1
/transl_table=1

/product="Csgd protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="MFNEVASHSHGHTLLITRKPSLQATALLQHLKQSLATIGKLLNIQ
RSLSDISAGCIVLMDMDADKLIHYQDNLSRNNNIKTLILNTPDQPYRREIENP
HINQVFATRDQEHVUSGLQILRGCEYFQKLASVYLITSGNVRVSTESALLTHRE
KELINKLRICASNNETIRSLFISENTYKTHLYNLFKLIAYAKRTQAVSMANDILRR"
gene
complement(2869..2896)
/gene="csgDERg"
complement(2869..2874)
/gene="csgDERg"
complement(2891..2896)
/gene="csgDERg"
3344..3376
/gene="csgBA"
3344..3349
/gene="csgBA"
3371..3376
/gene="csgBA"
3469..3924
/gene="csgB"
3469..3924
/gene="csgB"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Csgb protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKRKLPMULTIIGAGIATATVYDLARSEYFVAVNELSKSSFN
QAALIGQVGDNDARVROEGSKLSVTSQGGNNRAVDQAGNVPAYIQTGNANDA
SISQSAQNSAALIIQKSGKNAVITGYGTQKAVVQKQSHMAIRVQR"
3469..3531
/gene="csbB"
3532..3921
/gene="csbB"
/product="CsgB protein"
/function="nucleator"
3966..4421
/gene="csGA"
3966..4421
/gene="csGA"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="CsgA protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/db_xref="SWISS-PROT:P55225"
/translation="MKLIKVAAPRAIIVSSALAGVYPWQGGGNNHNGGSSGPDST
LSIYQGSANAALQSDARKSETTIVQSGNGADVGQADNSTIELTQNGFNNAT
IDQWNAQNSDITIVQYGNNAALVNQTAQSDSSVMVROVFGNNAATANQY"
3966..4025
/gene="csGA"
4026..4418
/gene="csGA"
/product="CsgA protein"
/function="fiber subunit"
4483..4809
/gene="csGC"
4483..4809
/gene="csGC"
/codon_start=1
/transl_table=1
/product="CsgC protein"
/protein_id="CA05318.1"

AUTHORS	McClalland, M., Sanderson, K.E., Spiehl, J., Clifton, S.W., Latifille, P., Courtney, L., Powell, S., Ali, J., Dante, M., Du, F., Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A., Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Florea, L., Miller, W., Sorenking, T., Nhan, M., Waterson, R., and Wilson, R.K.
TITLE	Complete genome sequence of <i>Salmonella enterica</i> serovar Typhimurium 172
JOURNAL	Nature 413 (6858), 852-856 (2001)
MEDLINE	21534948
PUBMED	11677609
REFERENCE	2 (bases 1 to 22411)
AUTHORS	.
CONSIST	Direct Submission
JOURNAL	Submitted (25-MAR-2001) Genome Sequencing Center, Department of Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT	Supported by NIH grant 5U 01 AI47393

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>.
and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.org>
<http://ecocyc.Pangeasystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and RegunDB; http://kinich.cifn.unam.mx:8850/bp/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

1. 22411

434. .130

434.439

454. .130

CDS

/gene="STM1127"
complement(1414..2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=11
/product="putative transcriptional regulator"
/protein_id="AAI20058.1"
/db_xref="GI:16419643"
/translation="MEPPPRRLKPKILDTLGAQOKSLTRASORIAVYLLAPFQQVI
SSIALDRETKGAEATVAFRCRTLGKGFOPKNDIAELATTSDDSPLLDMVEVER
SDDAHAIIGLKQNTISNVLSSETNLIDMQOVLGVADLRSHSGYITFEVGSGITDAL
MKHKMLRGLDVAASNHNFMWQAOTLLKAGDVAVGSHSTSEPTVASHLAQADAQAA
TTVALITHNIGSPLCEADPFLCINGNRQGLQDSTGYTAKDLFVDLLYTLTVOSPEE
QARESKLTALNALMDTK"
complement(12298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB: STMS1H001399"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=11
/product="putative sodium/glucose cotransporter"
/protein_id="AAI20059.1"
/db_xref="GI:16419644"
/translation="MITHSFGLVNYLVLFEGYLLAMLVGVYFSRRKQTADDYFRGGGR
VGMAGAGSVPAFTLSITFMSIPAKAFTSDMTPIIQYLAILPLVFYFIPEPFKKR
LVTAYEVELLEARFPVRCRLTFMSNFMLEFHGRAIIITFLVALRPALIDPVILVLI
LISWCATITMMGIGIEGVETWDIOGLLGSATILIFVICLXKGIGDEIFTYTOOA
DKFPATQGHMSWTSTVPVLMIGFLPANIQOFTRASQDVVRIVYTBSIBETKTLLT
NKLAVAIVIFFFAIGSALFVYIQHPDLDPGEVNTGILPLFVTEPVGITACLIITA
ALFAAQSISSFLNSISSCFNSDIYORLSHKRKTPENRMKIATLVILVAGLSASSAS
RWLVNADEBEIWDAFNLSILGIMGPMTGIFMGIGFFKRANAGSAVLGIITSIVILVGA
RAVLTDLPFYPYIGLSLVISIGVIFAPLPAPPLTLDKEPKEPYTL"

RBS

/complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB: STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=11
/product="putative inner membrane protein"
/protein_id="AAI20060.1"
/db_xref="GI:16419645"
/translation="MSLARLEQSVENHGCHIVSCQVPYSGPMDKPEITVAAMAQAAS
ACAVALIRIGIENLTRVPHLSVPRIIGIKEDLNGSPVRIPTVODVADAQAADIITI
ADPADESPENDISLIRIRIHGLIAMDGSTNEGISCQKITEFGITLSTGYRPP
ITPVEDPLAMVTQLSHAGCRVTAESGRYTPPALANAIEBHGAAMVATVSALIRIBHTQ
WFSSHVKR"

BBS

complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB: STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
5599..6759
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=11

gene

CDS

Query Match	Best Local Similarity	Matches	454; Conservative	0; Mismatches	2; Indels	0; Gaps	0;
Qy	1	ATGAAAAACAATTGTTATTATGATGTTGACAACTACTGGGTGCGCTCGGATTGCAACC	60				
Db	17272	ATGAAAAACAATTGTTATTATGATGTTGACAACTACTGGGTGCGCTCGGATTGCAACC	17333				
Qy	61	GCGCAAAATTATGATCTGGCTCGTTGAGAAATATATATTTTGGGTAATGAAATTAACCAAG	120				
Db	17332	GCGCAAAATTATGATCTGGCTCGTTGAGAAATATATATTTTGGGTAATGAAATTAACCAAG	17399				
Qy	121	TCTTCATTATATGACGGCGGCATTATTTGTCAGATCGGCAACGATATATAGTGGCAGAGTA	180				
Db	17392	TCTTCATTATATGACGGCGGCATTATTTGTCAGATCGGCAACGATATATAGTGGCAGAGTA	17455				
Qy	181	CGCCAGGAGGATCAAAACTATTTGTCGCTATTATTCACAGAAAGAGAAATTAATCGGGCG	240				
Db	17452	CGCCAGGAGGATCAAAACTATTTGTCGCTATTATTCACAGAAAGAGAAATTAATCGGGCG	17511				
Qy	241	AAAGTCACACGACGAGGAAATTAATCTTTGGCTATATTTGAGCAACCGGCAATGCCAAC	300				
Db	17512	AAAGTCACACGACGAGGAAATTAATCTTTGGCTATATTTGAGCAACCGGCAATGCCAAC	17571				
Qy	301	GATGCCAGTATATCGCAAAAGCGCTTACGGTAATAGTCAGACTATTAATCCAGAAAGGTTCT	360				
Db	17572	GATGCCAGTATATCGCAAAAGCGCTTACGGTAATAGTCAGACTATTAATCCAGAAAGGTTCT	17633				
Qy	361	GGAATTAAGCCCAATATTAATCCAGTACGGTAACGAGAAACAGACAGTTGATGCGAATA	420				


```

/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/translation="MATSRLPIHPVKMRGIGLTHALADISGLAFDRIPMTESDGT
FITRQPPQWRPTSPPLHDLHLTPAGSSALVRPTMPTPODAPTEPWGHHPTARVA
PTAINQMLSGFSPRDVQICWGPOLTRVYKHNAVPLGFAQYPTILNLSKRDLOQ
KCPAGVQMEQFRPNLVSGVAMVEDSKVIRIGDIVIVDKPCSRCLFTTVSPKQ
KHPSGPLATLQAFRTAQDNGVDVQNLIRANGSVIRVGDVEHLIATAPAKAYATT
LDGVTPEKHDPDGVTTIDMOGQTFCCANNQVLEQLERQIRIPYSCRAGICCCRIR
LLEGVSPLKSGAMGDDSTILSCSCVPTALRLEN"
complement(4737..4955)
/gene="STY1081"
/notes="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
/gene="STY1082"
5924..8032
/gene="STY1082"
5924..8032
/notes="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY_HABIN (P44524, P43945) fasta
scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI, Faera hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82756"
/db_xref="SPTREMBL:Q82756"
/translation="MNSLFTASTARGLEBELLEKLGAVGCVQVGGVHFGQDTRLI
YQSLMSRLAIRILPMGECKYVDLDLDFVQAIINTEIENPGATFVHSGANDTI
RNSQYGAWKVDAIVDAFTKRLPRPNVDRSPDLRIIVMINKESAIALSDGDI
LRGYRDTGLAIPKETTAAATVMSRQPGRLDIPMGSGTLLIEMAMATDAPRL
HRGHWBSGNAQDBAIIQVYKAEQRTARKGLAIBSHFPGSDARVIRASRANK
RAGIGELTFEVKQVAQSLNPLPKPGYGTVISNPYSGRILSDPALIALHLRTMK
NOFGWMLSLFSASPDLDGLQLRADQPKAKNPLDCVQKNYHIAETLADSKPATVA
EDVANRLKINKLEKEMARQEGIECYRLYADLDEYNAVVRGDMAYIOYAPKATVA
DAKARORLFDIIATLSVLGIIPNKILVTKFREKQKGNQYOKSKSEPFLESEYNA
RLMNLDLYLDTGLFIDIRIARMLGEMSKKDLNLESTYSGASVYAGGAGANTT
VDMKRTYEMAEERNLKNGLSGAPARLIGADCLGEMRANQPLIPIPTPFSNSKR
MEESFDVQRDHALMKDKRLRKGGTTFMSNNRGRFMDLEGLAEGLTAEITQKT
LSPDFARNROHNCMLIRAA"
5927..7060
/gene="STY1082"
5927..7060
/notes="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
/gene="STY1082"
6506..6541
/notes="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/notes="Faera hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faera hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
overlap"
Query Match 98.9%; Score 451.2; DB 1; Length 254050;
Best Local Similarity 99.3%; Pred. No. 1,2e-123;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 61 GCGCAATATATGATCTGGCTCGTTCAGATATATTTTGGGTAAATGAATTAAGCAG 120
DB 88469 GCGCAATATATGATCTGGCTCGTTCAGATATATTTTGGGTAAATGAATTAAGCAG 88528
QY 121 TCTTCATTTATTCAGCGCGGCATTATTTGGTCAAGTCGCACGAGTAATAGTCCAGAGTA 180
DB 88529 TCTTCATTTATTCAGCGCGGCATTATTTGGTCAAGTCGCACGAGTAATAGTCCAGAGTA 88568
QY 181 CGCGAGAGAGATCAAACTATTTGCTCGTATTTTCACAGAGAGAGAAATATTCGGCGC 240
DB 88589 CGCGAGAGAGATCAAACTATTTGCTCGTATTTTCACAGAGAGAGAAATATTCGGCGC 88648
QY 241 AAATTCACAGAGAGAGAAATATTAATCTTGGCTGATATTCAGCAACCGGCAATGCCAC 300
DB 88649 AAATTCACAGAGAGAGAAATATTAATCTTGGCTGATATTCAGCAACCGGCAATGCCAC 88708
QY 301 GATCCAGATATATCGCAAGCGCTTACGATATATGTCAGATATTAATTCAGAAAGTTCT 360
DB 88709 GATCCAGATATATCGCAAGCGCTTACGATATATGTCAGATATTAATTCAGAAAGTTCT 88768
QY 361 GGAATTAAGCGCAATATTTATCCAGTACGTACGAGAGAGAGAGAGAGAGAGAGAGAG 420
DB 88769 GGAATTAAGCGCAATATTTATCCAGTACGTACGAGAGAGAGAGAGAGAGAGAGAGAG 88828
QY 421 CAGTCGATATNGCTATTTGCGGTACCCAGCGCTAA 456
DB 88829 CAGTCGATATNGCTATTTGCGGTACCCAGCGCTAA 88864

```

```

RESULT 5
LOCUS AE016840/c 301983 bp DNA linear BCT 21-MAR-2003
DEFINITION Salmonella enterica subsp. enterica serovar Typhi Ty2, section 7 of
16 of the complete genome.
ACCESSION AE016840 AE014613
VERSION AE016840.1 GI:29137797
KEYWORDS
SOURCE
ORGANISM
Salmonella enterica subsp. enterica serovar Typhi Ty2
Salmonella enterica subsp. enterica serovar Typhi Ty2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
Deng, W., Liou, S.-R., Plunkett III, G., Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
Comparative Genomics of Salmonella enterica Serovar Typhi Strains
Ty2 and CT18
J. Bacteriol. 185 (7), 2330-2337 (2003)
JOURNAL MEDLINE 12644504
PUBMED 22531367
REFERENCE
2 (bases 1 to 301983)
Deng, W., Liou, S.-R., Plunkett, G. III, Mayhew, G.F., Rose, D.J.,
Burland, V., Kodoyianni, V., Schwartz, D.C. and Blattner, F.R.,
Direct Submission
Submitted (25-SEP-2002) Laboratory of Genetics, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
source
1..301983
/organism="Salmonella enterica subsp. enterica serovar
Typhi Ty2"
/mol_type="genomic DNA"
/strain="Ty2"
/serovar="Typhi"
/sub_species="enterica"
/db_xref="taxon:209261"
/complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
/notes="corresponds to STY1225 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1

```


CDS complement(7922..9133)
 /gene="f1gE"
 /locus_tag="t1743"
 /note="corresponds to STY1216 from Accession AL513382:
 Salmonella typhi CT18"
 /codon_start=1
 /transl_table=11
 /product="flagellar hook protein f1gE"
 /protein_id="AO69367.1"
 /db_xref="GI:29137806"
 /translation="MSFSQAVSGNLAAATNLDVIGNNANSATYGFPSGTAAPDMFA
 GSKVGLGVKAVGITDFTDGTGTTNTRGLDVAISQNGFRLVDSNGSVFVRNGQFKL

Query Match 98.9%; Score 451.2; DB 1; Length 301983;
 Best Local Similarity 99.3%; Pred. No. 1.2e-124;
 Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACATGCGTCCCTGGATTGCAAC 60
 Db 37807 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACATGCGTCCCTGGATTGCAAC 37748

Qy 61 GCGACAATTATGATCTGCTGCTTCAAAATTAATTTTGGCGTAATGAATTAAGCAAG 120
 Db 37747 GCGACAATTATGATCTGCTGCTTCAAAATTAATTTTGGCGTAATGAATTAAGCAAG 37688

Qy 121 TCTTCATTATACAGCGCGCCATTATGTCAGTCAGTCGCAAGCAATTAATGCGCAGATA 180
 Db 37687 TCTTCATTATACAGCGCGCCATTATGTCAGTCAGTCGCAAGCAATTAATGCGCAGATA 37628

Qy 181 CGCCAGAAAGATCAAACTATTCCTGCTTATTTTCAAGAAGAGAAATATATGCGGCG 240
 Db 37627 CGCCAGAAAGATCAAACTATTCCTGCTTATTTTCAAGAAGAGAAATATATGCGGCG 37568

Qy 241 AAAGTCAGCCAGCGGGAATTATTAATCTTGGCGTATTTGAGCAAAACGCGCAATGCCAAC 300
 Db 37567 AAAGTCAGCCAGCGGGAATTATTAATCTTGGCGTATTTGAGCAAAACGCGCAATGCCAAC 37508

Qy 301 GATGCCATATATGCGCAAGCGCTTAAGCTAATAGTGAAGTATATTCACAAAAGTCTCT 360
 Db 37507 GATGCCATATATGCGCAAGCGCTTAAGCTAATAGTGAAGTATATTCACAAAAGTCTCT 37448

Qy 361 GGAATAAAGCCCAATATTATCCAGTACCGTACGCAAGAAACAGCAAGTTAGTGCAGAAA 420
 Db 37447 GGAATAAAGCCCAATATTATCCAGTACCGTACGCAAGAAACAGCAAGTTAGTGCAGAAA 37388

Qy 421 CAGTCGATATGCTGCTATTCGCGTCAACCAACGCTTAA 456
 Db 37387 CAGTCGATATGCTGCTATTCGCGTCAACCAACGCTTAA 37352

RESULT 6
 STAGFA 1048 bp DNA linear BCT 26-JAN-1998
 LOCUS Salmonella typhimurium agfB and agfA genes.
 DEFINITION AJ000514.1 GI:2275119
 ACCESSION AJ000514.1
 VERSION agfA gene; agfB gene.
 KEYWORDS Salmonella typhimurium
 SOURCE Salmonella typhimurium
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.

REFERENCE
 AUTHORS Sukupolvi S., Lorenz R.G., Gordon J.I., Bian Z., Pfeifer J.D.,
 TITLE Notmark S.J. and Rhen M.
 JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
 MEDLINE 98053981
 PUBMED 9393832
 REFERENCE 2 (bases 1 to 1048)
 AUTHORS Sukupolvi S.S.
 TITLE Direct Submission

JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
 University of Turku, Kiinamyllynkatu, 20520, FINLAND
 FEATURES Location/Qualifiers
 source 1..1048
 /organism="Salmonella typhimurium"
 /mol_type="Genomic DNA"
 /db_xref="taxon:602"
 14..553
 /gene="agfB"
 14..553
 /gene="agfB"
 CDS 14..553
 /gene="agfB"
 /codon_start=1
 /transl_table=11
 /product="AgfB protein"
 /protein_id="CA04150.1"
 /db_xref="GI:2275120"
 /db_xref="SPTREMBL:O33801"
 /translation="MKRKLPMFLTIIEPIATATNYDLARSEYNPAVELSKSPNG
 AAIIGVGTNSARVREHSGSKLSVISOEGGNRRARVDAGNPNPVIETOTGNADA
 SISQAYGNNAISAIITOKSGSKNITQYGTQKTAIVVQKSHMALIQANITQYGTQ
 KTAIVVQKSHMALIRVQR"
 593..1048
 /gene="agfA"
 593..1048
 /gene="agfA"
 /transl_table=1
 /codon_start=1
 /transl_table=11
 /product="AgfA protein"
 /protein_id="CA04151.1"
 /db_xref="GI:2275121"
 /db_xref="SPTREMBL:O33802"
 /translation="MKLKLPAAPAIIVSGSNAVAVPQWGGGNNHNGNSSPDST
 LSIYQSSANAALALQSDARKSETTITQSGYSGADVGGADNDSTIELTQGFNNAT
 IDQNNANSDITVQYQGNNAALVNQTPASDSVWVRQYGNNAIPANQYN"

ORIGIN
 Query Match 83.9%; Score 382.8; DB 1; Length 1048;
 Best Local Similarity 94.6%; Pred. No. 1.3e-101;
 Matches 435; Conservative 0; Mismatches 7; Indels 18; Gaps 3;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACATGCGTCCCTGGATTGCAAC 60
 Db 14 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACATGCGTCCCTGGATTGCAAC 70

Qy 61 GCGACAATTATGATCTGCTGCTTCAAAATTAATTTTGGCGTAATGAATTAAGCAAG 120
 Db 71 GCGACAATTATGATCTGCTGCTTCAAAATTAATTTTGGCGTAATGAATTAAGCAAG 130

Qy 121 TCTTCATTATACAGCGCGCCATTATGTCAGTCAGTCGCAAGCAATTAATGCGCAGATA 180
 Db 131 TCTTCATTATACAGCGCGCCATTATGTCAGTCAGTCGCAAGCAATTAATGCGCAGATA 190

Qy 181 CGCCAGAAAGATCAAACTATTCCTGCTTATTTTCAAGAAGAAAGCAATTAATGCGG 237
 Db 191 CGCCAGAAAGATCAAACTATTCCTGCTTATTTTCAAGAAGAAAGCAATTAATGCGG 250

Qy 238 GCGAAAGTCAACGCGCAAGCAATTAATTAATCTTGGCGTATTTGAGCAAAACGCGCAATGCC 297
 Db 251 GCGAAAGTCAACGCGCAAGCAATTAATTAATCTTGGCGTATTTGAGCAAAACGCGCAATGCC 310

Qy 298 AACGATCCAGTATATGCGCAAGCGCTTACGGTAA-----TAGCGAGCTATT 345
 Db 311 AACGATCCAGTATATGCGCAAGCGCTTACGGTAA-----TAGCGAGCTATT 370

Qy 346 ATCCAGAAAGTTCTGGAATTAAGCCCAATTAATTAATCCAGTACGTTACGCAAGAAACAGCA 405
 Db 371 ATCCAGAAAGTTCTGGAATTAAGCCCAATTAATTAATCCAGTACGTTACGCAAGAAACAGCA 430

Qy 406 GTTGTAGTCAAGAAACAGTGCATATGCTATTGCGCTCA 445
 Db 431 GTTGTAGTCAAGAAACAGTGCATATGCTATTGCGCTCA 470

FEATURES	source
RESULT 7	
LOCUS	CSP515700
DEFINITION	Citrobacter sp. Fec2 csf gene, csf gene and csf gene.
VERSION	AJ151700
KEYWORDS	AJ151700.1 GI:31790491
SOURCE	csf gene; csf gene; csf gene; curlin-csf protein; nucleation
ORGANISM	Citrobacter sp. Fec2
REFERENCE	Citrobacter sp. Fec2
AUTHORS	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
TITLE	1 Enterobacteriaceae; Citrobacter.
JOURNAL	Zogaj, X., Bokranz, M., Nimtz, M., and Romling, U.
REFERENCE	Production of Cellulose and Curli Fimbriae by Members of the Family
AUTHORS	Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
TITLE	Infect. Immun. 72 (7), 4151-4158 (2003)
JOURNAL	2 (baes 1 to 2889)
REFERENCE	Romling, U.
AUTHORS	Direct Submission
TITLE	Submitted (11-NOV-2002) Romling U., Microbiology and Tumorigenesis
JOURNAL	Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
REFERENCE	1. 2889
AUTHORS	Location/Qualifiers

gene
complement(212. .862)
/gene="cs9d"
complement(212. .862)
/gene="cs9d"
/function="necessary for transcription of the cs9a operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNFHSIGHITLLITKPSLQATALLQHLKQSLASGLQHNIT
RLSDIISSCIVLVNMEADRKLIHYQONLGRKNNIKTLLITPDYGRINEMW
KINGFYVAADDEREVNGLGILRGECYSQSLASYLTTHSGNRYNSSSALLTIR
KEILNKRIIGASNIETIARSLFISENTVKTLLINLPKILAVKNQVAVSWANDLIR"
1623. .2078
/gene="cs9b"
1623. .2078
/gene="cs9b"
/function="nucleator"
/codon_start=1
/transl_table=1
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="FMMLITVLGAPGIASATSYDLASSEYNFAVELSKSSFN
QAAITQGVNNSAKNRQGSGLTSVSGEGSSNRPKADVQSGAFVAVIYAGSHNSDA
SISQNSNNTAIQKGSKANRITQYGTQKTAIVVQROSQMAIRVQR"
2119. .2571
/gene="cs9a"
2119. .2571

ORIGIN

Query Match	76.5%	Score 348.8;	DB 1;	Length 2889;
Best Local Similarity	85.3%	Pred. NO. 1.4e-91;		
Matches 389;	Conservative	0. Mismatch		

Qy	1	ATGAAAAACAATTGTTATTATTAATGATGTCACAACTAGGGGCGCCCTGGGAATTGCAAC	60
Db	1623	ATATAAAAACAATTGTTATTATTAATGATGTTACAGTACTGGGCGCCCTGGAAATTGCTTC	1682
Qy	61	GCACAAATTAATGATCTGGCTCGTTCAGAAATATAATTTTGGGTAAATGAATTAAGAAC	120
Db	1683	GCACAAGTTATGATTTAGCGATTTCCAAATATACTTGGCGGTAAATGAATTAAGTAAG	1742
Qy	121	TCTTCATTTAATCAGCGCGCCATTATTTGTCATCGGCACGGATTAATGTCACAGTA	180
Db	1743	TCTTCATTTAATCAGCGCGCCATTATTTGTCATCGGCAGAAATATACTGCAAAAGATG	1802
Qy	181	CGCCAGAAAGATCAAAAATACTTGTCCGTTATTTCAACAAGAGAGAAATAATCGGGCG	240
Db	1803	CGCCAGAGGGCTTAATACTTTGTGGTAGTTTCAACAAGAGGTGAAAGTAAACCGGGCC	1862
Qy	241	AAATTCACACCGCAGGGAATTAATACTTGTGCTATATTAGAGAAACGGGCATGCAAC	300
Db	1863	AAATTTACCAATCAGAGCCTTAATTAATTTTGGTATATTGCTCAGTGGGTCAATTCAT	1922
Qy	301	GATGCCAGTATATCGCAAAAGCGTTACGGTAATAGTGCAGCTATTAATCAGAAAGTTCT	360
Db	1923	GATGCCAGTATATCGCAAAAGTAATTAACGGTAATTAACTCGCATGATTAATCAGAAAGTTCT	1982
Qy	361	GGAAATTAAGCCCAATATTACCAAGTACGGTACGCAAAAACGACGTTGTGTGACAAA	420
Db	1983	GGAAATTAAGCAAAATTAATCAGATGAGTACGCAAAAACGACGTTGTGTGACAGAA	2042
Qy	421	CAGTCGATATGCTATTTCGGTCAACCCCAAGCTTA	456
Db	2043	CAGTCGAAATGGCAATTCGGCTTAATCAAGCTTA	2078

RESULT 8	AE015131	10370 bp	DNA	linear	BCT 18-OCT-2002
LOCUS	AE015131				
DEFINITION	Shigella flexneri 2a str. 301 genome.				
ACCESSION	AE015131	AE005674			
VERSION	AE015131.1	GI:24051313			
KEYWORDS					
SOURCE	Shigella flexneri 2a str. 301				

ORGANISM
Shigella flexneri 2a str. 301
Shigella flexneri 2a str. 301
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
1 (bases 1 to 1070)

TITLE	Author	Journal	Year
Genome sequence of <i>Shigella flexneri</i> 2a: insights into pathogenicity through comparison with genomes of <i>Escherichia coli</i> K12 and O157	Strom, J.M., Hou, Y.D. and Yu, J.	Journal of Bacteriology	2005

JOURNAL Nucleic Acids Res. 30 (20), 4432-4441 (2002)
 PUBMED 12384590
 REFERENCE 2 (bases 1 to 10370)
 AUTHORS Jin, O., Shen, Y., Wang, T., et al.

TITLE Wang, P.-Y., Wen, Y.-M., and Hou, Y.-D.
JOURNAL Direct Submission
Submitted (08-MAY-2003) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China

```

FEATURES
  source
    Location/Qualifiers
      1..10370
        /organism="Shigella flexneri 2a str. 301"

```



```
gene      /mol_type="genomic DNA"
          /strain="301"
          /setype="2a"
          /db_xref="taxon:198214"
          /complement(127..214)
          /product="cRNA-Ser"
          /note="anticodon: GGA"
          410..1387
          /gene="ycdW"
          /locus_tag="SF1027"
          /pseudo
          410..1387
          /gene="ycdW"
          /locus_tag="SF1027"
          /note="similar to Escherichia coli K12 putative
          dehydrogenase gi: 1787270 (326 aa). BLAST with identity of
          99% in 326 aa. This CDS contains an in-frame stop codon.
          The sequence has been checked and is believed to be
          correct."
          /pseudo
          /codon_start=1
          /transl_table=11
          1442..2179
          /gene="ycdX"
          /locus_tag="SF1028"
          1442..2179
          /gene="ycdX"
          /locus_tag="SF1028"
          /note="Residues 1 to 245 of 245 are 99 pct identical to
          residues 1 to 245 of a 245 aa protein from Escherichia
          coli O157:H7 EDL933 ref: NP_287168.1"
          /codon_start=1
          /transl_table=11
          /product="orf, conserved hypothetical protein"
          /protein_id="AAN42652.1"
          /db_xref="GI:24051314"
          /translation="MYPDLMHTVASTHAVSTLSDYIAOKKGKIKLPAITDGPDM
          EDAPHMHFIMRMIRMPRVGVIGLIGLEANIKNVDEGIDCGKMPDSDLIIGFHE
          FVPAPHDKATNTQAMISTIASGNVHITISHPGNPKETIDYKVAEAKAHQVALIENNS
          FPLHSRKGSEDCRAVAANVADAGGVVLSGSDSHATFTMGFEELCKILDVDFPLER
          ILNVSPRLNLFBSRGMAPIAEFADL"
          2203..2757
          /gene="ycdY"
          /locus_tag="SF1029"
          2203..2757
          /gene="ycdY"
          /locus_tag="SF1029"
          /note="Residues 1 to 184 of 184 are 99 pct identical to
          residues 1 to 184 of a 184 aa protein from Escherichia
          coli O157:H7 EDL933 ref: NP_287169.1"
          /codon_start=1
          /transl_table=11
          /product="putative oxidoreductase component"
          /protein_id="AAN42653.1"
          /db_xref="GI:24051315"
          /translation="WNERSTICRVIGSGSYRQPODPLVPIFLIRBKLANMPLEQ
          DELFLRLOKSCMTQVSNADYNALFTIGDECAVPYRSANVEDATEBAVAFISEGMLL
          ADTPADHIGTLILASWLEQSTEDSESLBTLSEVYLLPWCAGFLGKVEAHATTPFW
          RTMAPLTRDAISLAWDELDEBDEE"
          2811..3350
          /gene="ycdZ"
          /locus_tag="SF1030"
          2811..3350
          /gene="ycdZ"
          /locus_tag="SF1030"
          /note="Residues 1 to 179 of 179 are 98 pct identical to
          residues 1 to 179 of a 179 aa protein from Escherichia
          coli K12 ref: NP_415554.1"
          /codon_start=1
          /transl_table=11
          /product="orf, conserved hypothetical protein"
          /protein_id="AAN42654.1"
          /db_xref="GI:24051316"
```

```
gene      /translation="MCSISFMAAFSAIRMGNNILSTATTGIIISGKGVAVSIL
          SNAFLGCTATFACPDGGLKGLAISATLISGVYMANVITIGSLAPLLETICVITIG
          IAAFLKCTIAQKQLISFVPGTFIGACATFAGQGWKVLPSLALGLVGVYMKNSGLW
          IAAFSKCTAHRBOBKNA"
          complement(3414..4189)
          /gene="csqG"
          /locus_tag="SF1031"
          /pseudo
          complement(3414..4189)
          /gene="csqG"
          /locus_tag="SF1031"
          /note="similar to Escherichia coli K12 curli production
          assembly-transport component, 2nd curli operon gi: 1787274
          (278 aa). BLAST with identity of 97% in 257 aa. This CDS
          contains an in-frame stop codon. The sequence has been
          checked and is believed to be correct."
          /pseudo
          /codon_start=1
          /transl_table=11
          complement(4454..5662)
          /locus_tag="SF1032"
          complement(4454..5662)
          /locus_tag="SF1032"
          /note="Residues 1 to 402 of 402 are 100 pct identical to
          residues 1 to 402 of a 402 aa protein from plasmid R100
          ref: NP_052925.1"
          /codon_start=1
          /transl_table=11
          /product="transposase of Tn10"
          /protein_id="AAN42655.1"
          /db_xref="GI:24051317"
          /translation="MCSLDILHSDLYPCFPELHKRINSUTLACHALLDCTVLTTEL
          GRNLPYARTKNTKRTIDRLGNRHLKRLAVTRMASFTCSGNTMPTIVADSDIR
          EOKRLMWLRASVALHGWSVTLYERAFPLSEQCSKAKHQFLADLASILPSNTPLIYS
          DAGKXVPMYKSVKELGWMLSRVGVKYQVADLGAENKPISNLHDMSSSHKTYGKR
          LTKSNPISCOLILYKSRSGKGNORSTRTHCHHSPTIYSASAKEPMTLATNLPVEIR
          TPKNLVNISKRMQIESTFPDLKSPAYGLGRHRTSSSEFDMILILAMLOLTOML
          AGVIAQOGMDKHKQANTVRRNVLSVRLGMEVLRHSGYITTEDSDLSVAATLTLONL
          FTHGYVIGKL"
          complement(5995..6384)
          /gene="csqB"
          /locus_tag="SF1033"
          complement(5995..6384)
          /gene="csqB"
          /locus_tag="SF1033"
          /note="Residues 1 to 129 of 129 are 98 pct identical to
          residues 1 to 129 of a 129 aa protein from Escherichia
          coli O157:H7 EDL933 ref: NP_287173.1"
          /codon_start=1
          /transl_table=11
          /product="curli production assembly-transport component,
          2nd curli operon"
          /protein_id="AAN42656.1"
          /db_xref="GI:24051318"
          /translation="MKRYLRMIVAAEFPAAGNLHAEVVEVGLTDTHTVSIGDPY
          RASDDKESDYTGULTINERPSAKWSGMITTVQDVIFQTFPLPLKRSKTYVVAL
          IQTEBALNRQDINQALSTDLDLADER"
          complement(6389..7069)
          /gene="csqD"
          /locus_tag="SF1034"
          complement(6389..7069)
          /gene="csqD"
          /locus_tag="SF1034"
          /note="Residues 11 to 226 of 226 are 100 pct identical to
          residues 1 to 216 of a 216 aa protein from Escherichia
          coli O157:H7 EDL933 ref: NP_287174.1"
          /codon_start=1
          /transl_table=11
          /product="putative 2-component transcriptional regulator
          for 2nd curli operon"
          /protein_id="AAN42657.1"
          /db_xref="GI:24051319"
          /translation="MRSIKKMGFMFNEVHSIHGHTLLITKPSLQATLALQHLKQSL
```

ATGKTHNIGRSLDSSGSIILLDMWADKLIHYMODTLKRNKNNIKILLNTPED
 VYPDIEHMPHINGVFYAMEDQERVVAGVGGVTRGECYCTOKLASVLTTHSNTRVNS
 TESALTREREKEILINKLRIGASNNEIARSLFISENIVKHLINLFRKIAVKNRTOVNS
 WANDNLRR"

gene
 7766..8248
 /gene="csa8"
 /locus_tag="SF1035"
 7766..8248
 /gene="csa8"
 /locus_tag="SF1035"
 /note="Residues 10 to 160 of 160 are 99 pct identical to
 coli O157:H7 EDL933 ref: NP_287175.1"
 /codon_start=1
 /transl_table=11
 /product="minor curliin subunit precursor, similar to CsgA"
 /db_xref="GI:24051320"
 /translation="MYDOVQDNNKMKLLFMWLTILGAPGIAAAGVGLANSRYNFAV
 NEIKSSSENOAIIIGAGTNNISACLRGGSGTLAVVORSSNRKATIDGTQDYNLAVI
 DQAGSANDASISOGAVGNTAMIIQGSNNKANTQYGTQKTAVVOROSQMAIRVYOR

CDS

Query Match 73.3%; Score 334.4; DB 1; Length 10370;
 Best Local Similarity 83.3%; Pred. No. 2.6e-87;
 Matches 380; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATATATGATGTCACATATCTGGTGGCGGATTCGACAC 60
 DB 7793 ATGAAAAACAATTTGTTATATGATGTCACATATCTGGTGGCGGATTCGACAC 60
 QY 61 GCGACAAATTAATGATCTGGCTGCTTCAAGATTAATTTGCGGTAATGAATTAAGCAAG 120
 DB 7853 GCACAGAGTTATGATTTAGCTAATTCAGATTAATCTGCGGTAAATGATTAAGTAAAG 120
 QY 121 TCTTCATTATATCAGGGGCGCATTTATGTCAGATGTCGACGCGATATAGTCCAGAGTA 180
 DB 7913 TCTTCATTATATCAGGGGCGCATTTATGTCAGATGTCGCGGTAAATGATTAAGTAAAG 180
 QY 181 GCGCAGAGATCAAAATATGTCGTTATTTTCAAGAAAGAGAAATATTCGGGCG 240
 DB 7973 GCGCAGAGAGGCTCAAAATTTGGCGGTTGTCGCAAGAGTAAATGATTCGGGCA 240
 QY 241 AAATGCAACGAGGCGAGAAATTAATCTTGGATATTTAGCAACGCGCAATGCCAAC 300
 DB 8033 AAATGCAACGAGGCGAGAAATTAATCTTGGATATTTAGCAACGCGGCGCATGCCAAC 300
 QY 301 GATCCAGATATTCGCAAGCGCTTACGATATAGTGCAGTATTTATCCGAAAGGTTCT 360
 DB 8093 GATGCCAGATTTCCCAAGGCTTATGTAATCTGCGATTAATCCGAAAGGTTCT 360
 QY 361 GGAATTAAGGCGCAATTAATCCAGTACGTAACGAGAAACAGCAGTTGTAAGCAGAA 420
 DB 8153 GGTATTAAGCAATTAATTAACAGTATGTAATCAAAAAACGCGCAGTTGTAAGCAGAA 420
 QY 421 CAGTGCATATGCGCTATTCGCGTCAACCAAGGCTAA 456
 DB 8213 CAGTGCATATGCGCTATTCGCGTCAACCAAGGCTAA 456
 DB 8213 CAGTGCATATGCGCTATTCGCGTCAACCAAGGCTAA 8213

RESULT 9
 LOCUS ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
 DEFINITION E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
 ACCESSION X80754.1 GI:1147558
 VERSION X80754.1 GI:1147558
 KEYWORDS csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
 orfC gene.
 SOURCE Escherichia coli
 ORGANISM Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.
 REFERENCE 1 Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.
 AUTHORS

TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL

Expression of two csg operons is required for production of
 fibronectin- and congo red-binding curli polymers in Escherichia
 coli K-12
 Mol. Microbiol. 18 (4), 661-670 (1995)
 5641468
 8817489
 2 (bases 1 to 4680)
 Hammar, M.
 Direct Submission
 Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet
 Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
 SWEDEN

FEATURES
 source
 Location/Qualifiers
 1..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="23.15 minutes"
 1..3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement(133..966)
 /gene="csgG"
 complement(133..966)
 /gene="csgG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CA62277.1"
 /db_xref="GI:1147559"
 /db_xref="GOA:P52103"
 /db_xref="SWISS-PROT:P52103"
 /translation="MQRFLVAVMILSGCLTAPPKEARPTLMRAQSYKDLTHPA
 PKRTFVSQVYNQDETGQKPYAPSNFSTAVPOSATMLATKDSHFIPLEKQGLQ
 IAGDTQYQDQIAVNLRVNVSTGEILSSVNTSTTISYVQGVFRFIDYORLLGE
 VGTSENPEVMLCLMSAIEITGVIFILINDIRGLMDLONKARQNDIIVKRYHMSVPE
 S"

source

repeat_region

gene

CDS

gene

CDS

gene
 CDS

Location/Qualifiers
 1..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map="23.15 minutes"
 1..3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement(133..966)
 /gene="csgG"
 complement(133..966)
 /gene="csgG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CA62279.1"
 /db_xref="GI:1147561"

Query Match	73.0%;	Score 332.8;	DB 1;	Length 4680;
Best Local Similarity	83.1%;	Pred. No. 7.5e-87;		
Matches 379; Conservative	0;	Mismatches 77;	Indels 0;	Gaps 0;

RESULT 10	
LOCUS	AE000205
DEFINITION	AE000205 10346 bp DNA linear BCT 01-DEC-2000
ACCESSION	Genome.
VERSION	AE000205 U00096
KEYWORDS	AE000205.1 GI:1787265
SOURCE	
ORGANISM	Escherichia coli K12
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
AUTHORS	1 (bases 1 to 10346)
	Blattner,F.R., Plunkett,G. II, Bloch,C.A., Berna,N.T., Burland,V.,
	Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
	Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
	Mau,B. and Shao,Y.
	The complete genome sequence of Escherichia coli K-12
TITLE	Science 277 (5311), 1453-1474 (1997)
JOURNAL	9742617
MEDLINE	9278503
REFERENCE	2 (bases 1 to 10346)
AUTHORS	Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (16-JUN-1997) Guy Plunkett II, Laboratory of Genetics,
	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
	Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
	608-263-7459
REFERENCE	3 (bases 1 to 10346)
AUTHORS	Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
	University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
	Email: ecol1@genetics.wisc.edu Phone: 608-262-2534 Fax:
	608-263-7459
REFERENCE	4 (bases 1 to 10346)
AUTHORS	Plunkett,G. III.
TITLE	Direct Submission
JOURNAL	Submitted (13-OCT-1998) Laboratory of Genetics, University of

COMMENT

Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markborov@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Belyin. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

```

source          location/Qualifiers
                1..10346
                /organism="Escherichia coli K12"
                /mol_type="genomic DNA"
                /strain="K12"
                /sub_strain="MG1655"
                /db_xref="taxon:83333"
                /note="133"
                /codon_start=1
                /product="orf, hypothetical protein"
                /protein_id="AAC74113.1"
                /translation="MEDGVQMRDYLAGLDIASPEHOVMNTAKSEVAPSIKENTIS
                LHLTHYKMTESLDTFLNMPYPAFLFISNKPYLGNMLNDPAGVDQORVMCRFAFSD
                AKVFR"
                275..299
                /gene="b1028"
                /note="factor Sigma32; predicted +1 start at 1094963"
                389..1375
                /gene="ycdU"
                /note="synonym: b1029"
                389..1375
                /gene="ycdU"
                /function="orf; Unknown"
                /note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
                to 191 residues of an approx. 288 aa protein Y320_MWGE
                SW: P47562"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC74114.1"

```

```

protein_bind
/gene="ycdU"
/note="central position to predicted promoter: -104"
/bound_moiety="Deor predicted site"
1253..1280
/gene="ycdU"
/note="No predicted promoter"
/bound_moiety="Deor predicted site"
1337..1365
/gene="ycdU"
/note="factor Sigma70; predicted +1 start at 1096049"
1494..1745
/gene="b1030"
1494..1745
/gene="b1030"
/function="orf; Unknown"
/note="083; This 83 aa ORF is 25 pct identical (7 gaps)
to 81 residues of an approx. 616 aa protein ALBU_RABIT SW:
P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGRTTCVYTKKKEVLYELFFKYGEGIDSLRPGQPAHC
VHG.SNMLSPVDPBSGLIPVCAIHERKARTFVALL"
1513..1926
/gene="ycdV"
/note="synonym: b1031"
1513..1926
/gene="ycdV"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS37_YEAST SW:
P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSPYFRTSSILNAAVRGGLTRFAPRAGSPULATVCPT
GROLSTPGRGSPRCVQYTKKPEVLSYELFFRYGEGIDSLRPGQPAHCVHGL
SNMLSPVDPBSGLIPVCAIHERKARTFVALL"
complement(2111..2198)
/gene="bexX"
/note="synonym: b1032"
complement(2111..2198)
/gene="bexX"
/product="tRNA-Ser"
/note="anticodon: GGA; CG Site No. 17878"
2269..2296
/function="RNA; tRNA"
2269..2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338..2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355..2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393..3370
/gene="ycdW"
/note="synonym: b1033"
2393..3370
/gene="ycdW"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPRA_CORGL
SW: P45637"

```

```

/codon_start=1
/translate=1
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAQRFSGVSDIIIFYHPTEDTQWMIETAKAIPOARVRAN
KSGNDSADYALVWHPVEMLRGDLKAVPLAGVDSILSKLOHPMLPSVLEPR
LEDQMGOMOEAVSOVLHWRFRDQYRIONSSHMOPLPEYHEDTETIGAGVL
GSKVAOSIQTRFPRRCMSRTRKSPQSGRARELSAFISQCRVLINLPLPPEY
GIHQDLLEKLPDGAIVLNLKRGHVNVEDDLAALDSGRVGMALDVNRPPLPPEP
LMQHPRTITTHVAITRPAVEYISRTIAOLEKGRVCGQVDRARG"
3278..3308
/gene="ycdW"
/note="Factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="Factor Sigma70; predicted +1 start at 1098094"
3396..3423
promoter

Query Match      73.0%; Score 332.8; DB 1; Length 10346;
Best Local Similarity 83.1%; Pred. No. 7.7e-87;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACGGGTGCGCTGGATTGCAAC 60
Db 8497 ATGAAAAACAATTGTTATTTATGATGTTGACAAATACGGGTGCGCTGGATTGCAAC 8556
Qy 61 GCGACAATTATGATCTGCTGCTTCAAAATTTTGGCGTTAATGAAATTAAGACAG 120
Db 8557 GCACACAGTTATGATTTAGTTACAAATATTAATCTTCGGGTAAATGAAATTAAGTTAG 8616
Qy 121 TCTTCATTATTCAGCGCGCATTTATGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 180
Db 8617 TCTTCATTATTCAGCGCGCATTTATGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 8676
Qy 181 CGCAGAGAGATCAAACTATTTGCTGCTTATTCACAAAGAGAGAAATTAATCGGCG 240
Db 8677 CGCAGAGAGATCAAACTATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8736
Qy 241 AAAGTCACCGACGCGCAATTTATTAATCTTTCGCTATATTCAGCAACCGCGCAATTC 300
Db 8737 AAAGTCACCGACGCGCAATTTATTAATCTTTCGCTATATTCAGCAACCGCGCAATTC 8796
Qy 301 GATGCCAGTATATCGCAAGCGCTTACGCTATATGTCAGCTATATTCAGCAAGCGCTT 360
Db 8797 GATGCCAGTATATCGCAAGCGCTTATGCTATATGTCAGCTATATTCAGCAAGCGCTT 8856
Qy 361 GAAATTAAGCGCAATTTATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 420
Db 8857 GGTAAATTAAGCAATTTATTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 8916
Qy 421 CAGTCGCAATGCGCTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 456
Db 8917 CAGTCGCAATGCGCTATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8952

```

```

RESULT 11
LOCUS      D90741      15047 bp      DNA      linear      BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION  D90741 AB001340
VERSION     D90741.1 GI:1561509
KEYWORDS   Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
SOURCE      csgA; ycdW; cIs; nov; mdg.
ORGANISM   Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1
AUTHORS     Oshima, T., Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
            Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
            Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
            Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y.,
            Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K.,
            Wada, C., Yamamoto, Y., Yano, M. and Horiuchi, T.

```

```

TITLE      A 718-kb DNA sequence of the Escherichia coli K-12 genome
            corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL    DNA Res. 3 (3), 137-155 (1996)
MEDLINE    97061202
PUBMED     8905232
REFERENCE   2
AUTHORS     Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A., Horiuchi, T.,
            Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
            Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
            Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
            Motomura, K., Nakamura, Y., Nishio, Y., Oshima, T.,
            Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
            Yamamoto, Y. and Yano, M.
TITLE      The systematic sequencing of the Escherichia coli genome in Japan
JOURNAL    Unpublished
REFERENCE   3 (bases 1 to 15047)
AUTHORS     Mori, H.
TITLE      Direct Submission
JOURNAL    Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
            and Technology, Res. & Edu. Center for Genetic Info., 8916-5
            Takayama, Ikoma, Nara 630-01, Japan
            (E-mail: hmori@gtc.aisr-nara.ac.jp, Tel:81-7437-2-5660,
            Fax:81-7437-2-5669)
COMMENT     Collaboration Information:
            Project: The Japan E.coli genome DNA sequencing project
            Group: The Japan E.coli genome DNA sequencing group
            Members: (1995.4 - 1996.3)
            Alba, H., Baba, T., Fujita, K., Hayaishi, K., Honjo, A.,
            Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
            Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
            Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
            Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
            Nakamura, Y., Nishio, Y., Oshima, T., Saito, N.,
            Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
            Yamamoto, Y. and Yano, M.
            Headed by:
            Name: Takashi Horiuchi
            Address: National Institute of Basic Biology, Okazaki, 444, Japan
            E-mail: kihori@nibb.ac.jp
            Information operator:
            Name: Hirotsada Mori
            Address: NARA Institute of Science and Technology,
            Ikoma, 630-01, Japan
            E-mail: hmori@gtc.aisr-nara.ac.jp
            URL:
            The Japan E. coli genome database
            http://bsw3.aisr-nara.ac.jp.
            Location/Qualifiers
                1..15047
                /organism="Escherichia coli K12"
                /mol_type="genomic DNA"
                /strain="K12"
                /db_xref="taxon:83333"
                /map="23.7-24.0 min"
                /clone="Kohara clone #231"
                /note="Nucleotide position 1097101-1112147 from the
                initiation site of Thra (0 min.). ~This clone is from
                Kohara lambda miniset library."
                <1..276
                /note="unnamed protein product; ORF_ID:0230#5"
                /codon_start=1
                /transl_table=11
                /protein_id="BAJ5822.1"
                /db_xref="GI:4062605"
                /translation="DLPFSESNALSTFPAPKPLMLRLPILIVETWITLVSVALTIVE
                WRIFEAIVFLCTVPCILTYFYFCFLPHNDFMMACDMTFRMGHFNK"
                1152..2090
                /note="ORF_ID:0231#1
                similar to SwissProt Accession Number P45637"
                /codon_start=1

```

CDS

gene
CDS
/transl_table=11
/product="Hypothetical 33.0 kd protein in prob-proa
intergenic region."
/protein_id="BA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFYPFTDPTOMWIEALRAKIPQARVRAMKSGDNDSDADYALV
MHPVEMLAGRDLKAVFALGAGVDSTLSKLAHPMLNPSPLRLBEDTQGEQMOEY
AVSQVLMHFRPDPDYRIONSSHQPLPEYREDPTTIGIGAGVLSKVAQSLQVME
PLRCMSRTKSWPVGVSFAGREELSAFISQCRVLTINLPETVGIINQLEKLPD
GAILNLARGVHVVDLLAALDSGRVGAALDFNREPLPPEPSPLMQHBRVITTHV
AAITRAEAVEYISRTIAQLEKGERVCGQVDRANGY"
2145..2882
/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=11
/protein_id="BA35824.1"
/db_xref="GI:4062607"
/translation="MYPVDLMHTVASTHAYSTLSYIAQAKQKIGLPAITDHPDM
BDAPHHMFINKRIMPRVVDGVLIRGEIANIKVDEIDCSGMFSDLLIAGPHE
PVFAPHDKATNTQMIAITIASGVNHIISHGNPVEIDYKAVAEAAKQVALLKINS
SPLHSRKSSEDNCRVAAVADAGVVALGSDSHITFTMGEPEECLIKIDAVDFPPER
ILNVSFRLNLFLESRVAAPIAEFADL"
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=11
/product="Hypothetical protein H1543"
/protein_id="BA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVIGSLVYROPDPLVPLTLIREKLANMPLDQ
DELLTRLOKSCDWTQVADYNALFTDECAVPFISAVGEATEVAEVAFLSEGMFL
RTMAPLRTDASIMBDELBDESE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=11
/protein_id="BA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAAFSAIRGNILSLIATITGLISGIMGVAVSLGL
SVAGFLGCTAFACQGGKIGKLAISAAITLSGVVMAWIIYGSALPHLEILGYVIG
IYAFIMCTOAKOLLSFVGTGTGACATFAGCGDKLVPLSLALGILFGYAMNQSILM
LAARSAKTAHREOIKENKA"
complement(4117..4950)
/gene="cs9g"
/complement(4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgB precursor."
/protein_id="BA35827.1"
/db_xref="GI:4062610"
/translation="MORLFLVAVMLISGLTAPPEKARPTLMPAQSVDLTHLPA
PTGKIRVSVNIODETSGQRPYPASNFSTAVPOSATAMVTLKXSRFLEBQGO
NLNERKIRIAQENGIVAILNNRIPLQSTLAANIWEGSIIIEYENVSQGVARVIG
IGADTOYDQIDAVNIRVNVSTGELISVNTSKILLSEVQAGIFRTIDQRLDEE
VGYTSNEPVMLCMSAIEFGVIFLINDGIDRLMLQNKAEKQNDILVKYHMSVPE
S"
complement(4977..5393)
/gene="cs9g"
/complement(4977..5393)
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgF precursor."
/protein_id="BA35828.1"

gene
CDS
/db_xref="GI:4062611"
/translation="MRVHAVVLMILSPISNAGTMTFQERNFNGNPNNGAFILNS
AOANSYKDPNDPGLTETPSALNDNTQALQSLIGGLSLNINTQPGRMVNTDYIV
DIARNDQLOLVNDRKTKGTGTSTIQVSGLNNTDF"
complement(5418..5807)
/gene="cs9g"
/complement(5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=11
/product="Curli production assembly/transport component
CsgB precursor."
/protein_id="BA35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRMTVAEFLPAAGNLHAVEVFGLLTHTVSSIGHDFY
RASFDSKESDYGNLTINERPSARWGSMTTITTVQDVIFQFLPLKRDKEKTVFAL
IOTEEALNRQINQALISTGDLAHDFF"
complement(5812..6462)
/gene="cs9g"
/complement(5812..6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=11
/product="Probable csgB operon transcriptional regulatory
protein."
/protein_id="BA35830.1"
/db_xref="GI:4062613"
/translation="MFNEVSHIGHTLITKSSLOATLQHLQKSLAITGKLNIO
RSLDDISGSIILDMMEADKLIHWMDLTSKNNIKLILNTPEDEPYRDIENWP
HINGVFSRMEQDERVVGLOGVIRGEYFQKASVLTISGNYRNSYESALITRE
KEINLXLRIGASNNELRSFISENTVTKHLYLFRKIAVKNRTQAVSWANDLNR"
7217..7672
/gene="cs9g"
7217..7672

gene
CDS
Query Match 73.0%; Score 332.8; DB 1; Length 15047;
Best Local Similarity 83.1%; Pred. No. 7.8e-87;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
1 ATGAAAAACAATTTGTTATTTATGATGTGACAAATACGCGCTGGGATTCACAC
7217 ATGAAAAACAATTTGTTATTTATGATGTGACAAATACGCGCTGGGATTCACAC
61 GCGACAAATTTATGATCTGCTGCTCAGAAATATTTTGGGTAAATGAATTAAGCAAG
7277 GCGACAGTTATGATTTTACCTAATTCAGAAATATACCTTCGCGGTAAATGAATGAATGA
121 TCTTCATTTATCAGCGCGCCATTTATGTCAGTCCGACGATTAATAGCCAGATA
7337 TCTTCATTTATCAGCGCGCCATTTATGTCAGTCCGACGATTAATAGCCAGATA
181 CCGCAGGAAGATCAAACTATTGTCGTTATTTTCACAGAAAGAGAAATTAATCGCGC
7397 CCGCAGGAGGCTCAAACTTTTGGCGGTTTTCGCAAGAGGTATACCAACCGCGCA
241 AAGTCGACCGACGCGGGAATTAATTAATTTGCGTATTTAGCAAAACGCGCAATGCGCAC
7457 AAGATTGACAGACGAGGATTAATTAATTTGATTAATTAAGCGCGGAGTGGCAC
301 GATGCAATATATGCAAAAGCGCTTACGTAATTAATGTCACGCTTATTCAGAAAGGTTCT
7517 GATGCAATATTTGCAAGGCTTATGTAATTAATGTCAGTAATTAATTCAGAAAGGTTCT
361 GGAATTAAGGCAATATTAATCCAGTACGATACGCAAGAAACAGAGTTAGTACGAGAA
7577 GGTATTAAGCAATATTAATCAAGTATGTAATCAAAACGCAATTTAGTACGAGAA
421 CAGTCGATATGCTATTTGCGTACCGCAACGCTAA 456

DB 7637 CAGTGCMAATGGCTATTGCGTGAACAACGTTAA 7672

RESULT 12
AE016981
LOCUS Shigella flexneri 2a str. 2457T DNA linear BCT 22-APR-2003
DEFINITION genome.
ACCESSION AE016981 AE014073
VERSION AE016981.1 GI:30040616
KEYWORDS
SOURCE Shigella flexneri 2a str. 2457T
ORGANISM Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
REFERENCE 1 (bases 1 to 292504)
We1,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Complete Genome Sequence and Comparative Genomics of Shigella
flexneri Serotype 2a Strain 2457T
Infect. Immun. 71 (5), 2775-2786 (2003)
2 (bases 1 to 292504)
We1,J., Goldberg,M.B., Burland,V., Venkatesan,M.M., Deng,W.,
Fournier,G., Mayhew,G.F., Plunkett,G. III, Rose,D.J., Darling,A.,
Mau,B., Perna,N.T., Payne,S.M., Runyen-Janecky,L.J., Zhou,S.,
Schwartz,D.C. and Blattner,F.R.
Direct Submission
Submitted (13-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA

TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS

FEATURES
SOURCE
1. 292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement(223..441)
/gene="infA"
/locus_tag="S0884"
complement(223..441)
/gene="infA"
/locus_tag="S0884"
/function="factor; Proteins - translation and
modification"
/note="residues 1 to 72 of 72 are 100.00 pct identical to
residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
/codon_start=1
/transl_table=1
/product="protein chain initiation factor IF-1"
/protein_id="AAPI6348.1"
/db_xref="GI:30040616"
translation="MAKEDNIEMOQIVLETLPTMTFRVLENGHVVAHAISGKQRKQV
IRITGDKVTEIETPLYSKRIYFRSR"
complement(726..1430)
/gene="atf"
/locus_tag="S0885"
complement(726..1430)
/gene="atf"
/locus_tag="S0885"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA
modification"
/note="residues 1 to 234 of 234 are 100.00 pct identical
to residues 1 to 234 of 234 from Escherichia coli K-12 :
B0885"
/codon_start=1
/transl_table=1
/product="tycyl, phenylalanyl-tRNA-protein transferase"
/protein_id="AAPI6349.1"
/db_xref="GI:30040616"
translation="WRLIVQLSRHSIAFPSREGALREPNGLLALGSDLSPARLLAYQR

gene
CDS
complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/function="transport; Transport of small molecules; Other"
/note="residues 1 to 573 of 573 are 97.03 pct identical to
residues 1 to 573 of 573 from Escherichia coli K-12 :
B0886"
/codon_start=1
/transl_table=1
/product="ATP-binding component of cytochrome-related
transport"
/protein_id="AAPI6350.1"
/db_xref="GI:30040619"
translation="WRALPLALYKHKMMLSGIVLAVTLASIGLTLSCMPLS
ASAVAGVAGLYSENYMLPAAGVRCALITRAGRYFERLVSHDAFVRLQHLRIYFSK
LLPLSPAGLARVQSGELNVADVDTLDHLYRVLSPVLAQAFVIMVITIGLSLDF
TLAFTLGSVMLLTLPLMPPLPYRAGKSGQNLTLRGYRQQLTAMQSGAELTFIGA
SDRYRQLENTIEIOMLEAORROSETLTLASIMLIGALAVIILMMASGGVGNAPQ
GALIALVFCALAAFEKALAVYTAFCPLGVYASAVYISDITDQKEVTPPTOTRVA
DRSLVLRDVFPTPEOSQDLKITSIQVNAHEITAILRGRCCKSTLQQLTAMDP
QCGEITLNDSPILASALRQITISVPRFSAULRNLRLNLAPOSSDEALSEL
RRVLEKLELDEAGLNSWIGEGRQISGSELRRLIATLLADALVLDDETEGIDAT
TESQIILLAKMKREKTVLWYTHRLRGLSRFOQIIVMDNQIIRGTHAELLARQRY
YQPKQGI"
complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
/function="transport; Not classified"
/note="residues 1 to 588 of 588 are 92.85 pct identical to
residues 1 to 588 of 588 from Escherichia coli K-12 :
B0887"
/codon_start=1
/transl_table=1
/product="ATP-binding component of cytochrome-related
transport"
/protein_id="AAPI6351.1"
/db_xref="GI:30040620"
translation="MNSKROKELTRMLKQOSVISQWMLNISRLLGFSGLIIQAQWF
MARLOMIMENIPREALLPFTLVLTFLRAVWMLREBVGTHAGOHIRFAIRQV
LDRLOQAGPAMIOCKPAGSWATYVEQIDMDMDVYAVYLOMALAVSPVLLIYAIRP
SNVNAIILIGTAPLIPFLNALVGMGADNRRNRLYALRISGFILRLRMEYLRIF
GREGARIESIRASSEDPRQRTMEVRLAFLSSGILFEFTSLISIALVAVYGFSTGL
DPGHYDGVTLAAGFLALILAPEFQPLRDLGTFYHAKQAQVGAADSLKTFEMSPLAH
PQREAEALATDPLTIPABDLFTISPEKTLAGLNTLTPAGCAVAVVSGSGSSIL
LNLISGLFVQSGIRINGIELRDIISPESKRLSWGQONLPAAKLRDNLVLRPDA
SEQELQALDNNAVSEITPLRPGQVDPVCDQARLSVCAQARAVAVARALINPSLIL
LDEPASLDRHSRQVMEALNALSROTTLVTHQLDLADMDVITWMDSRITIEQGR
YAEISVAGSPATLLAHROEI"
complement(5083..6048)
/locus_tag="S0888"
complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/function="enzyme; 2-Deoxyribonucleotide metabolism"
/note="residues 1 to 321 of 321 are 95.95 pct identical to
residues 1 to 321 of 321 from Escherichia coli K-12 :
B0888"
/codon_start=1
/transl_table=1
/product="thioredoxin reductase"
/protein_id="AAPI6352.1"
/db_xref="GI:30040621"

```
/Locus_tag="S0889"  
6592. :7086
```

```
/Locus_tag="S0889"
```

regulator, central intermediary metabolism
Pool, multipurpose conversions
/note="residues 1 to 164 are 100 %
"

ESCHERICHIA COLI K-12

```

/coach_start=1
/transl_table=1
/product="translator for Java"

```

/protein_id="API16353.1"

```

/translation="MVDSSKRPCKDLRIDRINILNELQDKGRISNVELSKRVGLSP
LERVPI EPOCFCGWS"
uu_xref="GI:30040622"

```

QOECHLVSGDFDYLLKTRVPDMSAYRKLLGETLLRLPGVNDTRITYVMEEVKQSNRLV
KTR"

gene="ftsk"

```
221. .11249
gene="ftsK"
```

function="phenotyp

residues 1 to 737 of 1342 are 68.24 pct identical to residues 1 to 737 of 1329 from *Escherichia coli* K-12 : 890"

```
codon_start=1
transl_table=11
```

```
protein_id="AAP16354.1"
protein_xref="GI:30040333"
```

translation="MSQYEYTEDKEVTLTKLSSGRRLLEALLILVLPFAVLMALLSP
PSDDPSWQTAWHEPIHNLGMPGAWIADTLFEIEGYMAVETDITVFAVGLMALLSP

DEJEDYFANSRIRIGVALLILTS CGLA INADDIWYFASGGVIGSLSTLLOPLHSGEITIALLCWAGLTLFTGMSWVTIAELGGMILNILTFASNRTRRDDPTWDEDEYE

ITYTARGVAADPDDVLFSGNRTQPEYDEYDPLLSAPITEPVAVAAATATQSWA
 VERPTQTPPVAASVDVPPSPQPTAVMOBUDPCDNRAT
 VEEPTQTPPVAASVDVPPSPQPTAVMOBUDPCDNRAT

EEPLQQPVPQPPYAPAEQPAAQPPYAPAEQPVAGNMQAEEQSTFAPLSTYO
OTYQPAAGEPLYQOPQVPEQOPVEEPVEETKPARPPIYYEFEVEEKRAPE

AAAPVFSLNANGSGPRPEQVEKIGIQPLRPKRIRVPTRELASTYGIKLPSORAAEEKA
AORNROYDSGDQDYNMDETAKMOODT...
A

ELARQFAQTQQORYSGEOPAGANPFLSLDDFEFSPMKALLDDGPHBPLFTPIVEPVQ
QOPVAPQQOYQPBPQFVAPQQOYOOPQQQVABPPQYQNNBBPQAQBQVQVQVQVQVQ

LTPPSEVEVDTFEALOMARLVEARLADFRKADVNVNSPCVITREELNLAGY
QFQIQQQFQAVAPQPOYQQBPQPVPAPQPDTLHLPLMRNGDSRPLHKPTTLPBLSL
LRISNTSPDTADGCGMMAA

Score 332.8; DB 1; Length 203504
73.0%;

83.1%;	Pred. No. 8.6e-87;		
0;	Mismatches 77;	Indels 0;	Gaps 0;

AAATGTTATTATGATGTTGACATACTGGGTGCGCCTGGGATTGCAAC 60

AAATTGTTATTATGATGTTACATACTGGTGGCCCTGGGATTGCAGCC 21087

ATGATTTTACCTAATTTCATTTTTCGGTAATGAATTAGCAAG 120

ATCAGCGCCGCAATTATTGTCAAAGTCGGACCGATAATAGTCCCCCTCCG, 100

NOT WITHSTANDING THE FACT THAT

[illegible]

[illegible]

```

transposase [Shigella flexneri]
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAN79735.1"
/db_xref="GI:26107536"
/translation="MAERLGEIQKRVITVCRREADIMHYLHYKVSHGQRFVNTAONS

Query Match
Best Local Similarity 73.0%; Score 332.8; DB 1; Length 306358;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTGTTATTTATGATGTTCACAACTACGCGGTGCGCTGGATTGCAACC 60
DB 44773 ATGAAAAACAATTGTTATTTATGATGTTCACAACTACGCGGTGCGCTGGATTGCAACC 60
QY 61 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
DB 44833 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
QY 121 TCTTCATTTATATCAGCGCCATTTATGTCAGAGTGGCGAGATATATGTCAGAGTA 180
DB 44893 TCTTCATTTATATCAGCGCCATTTATGTCAGAGTGGCGAGATATATGTCAGAGTA 180
QY 181 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
DB 44953 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
QY 241 AAAGTCAGCAGGAGGAGGAAATTTATTAATCTTGGCTATATGACCAACCGGCATCCAC 300
DB 45013 AAAGTCAGCAGCAGGAGGAAATTTATTAATCTTGGCTATATGACCAACCGGCATCCAC 300
QY 301 GATGCCGATATTCGCAAGAGCGCTTACGTTAATGATGACGATTTATTCAGAAAGTTCT 360
DB 45073 GATGCCGATATTCGCAAGAGCGCTTACGTTAATGATGACGATTTATTCAGAAAGTTCT 360
QY 361 GGAATTAAGCCCAATTTACCACTACGTAAGTACGAGCAAGAAACGCGTTGTGTCAGAAA 420
DB 45133 GGAATTAAGCCCAATTTACCACTACGTAAGTACGAGCAAGAAACGCGTTGTGTCAGAAA 420
QY 421 CAGTGCATATGCTTATTCGCTGCTACCCACCGCTAA 456
DB 45193 CAGTGCATATGCTTATTCGCTGCTGCTGACACACGCTAA 456
QY 45193 CAGTGCATATGCTTATTCGCTGCTGCTGACACACGCTAA 45228

RESULT 14
AF275733 1711 bp DNA linear BCT 15-MAY-2001
LOCUS Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
DEFINITION CsgB protein (csgB) genes, complete cds.
ACCESSION AF275733
VERSION AF275733.1 GI:14039399
KEYWORDS Escherichia coli
SOURCE Escherichia coli
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1711)
AUTHORS Unlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Mutations in the csgD promoter associated with variations in curli
JOURNAL expression in certain strains of Escherichia coli O157:H7
MEDLINE Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
PUBMED 21218556
REFERENCE 2 (bases 1 to 1711)
AUTHORS Unlich,G.A., Keen,J.E. and Elder,R.O.
TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
RESEARCH Center, State Spur 180, Clay Center, NE 68933, USA
FEATURES
Location/Qualifiers
1..1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
source

```

```

/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757..1212
/gene="csgB"
757..1212
/gene="csgB"
757..1212
/codon_start=1
/gene="csgB"
/transl_table=11
/product="CsgB protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKKLLPMLITLAPGIAAAGYDLANSEYFANVELSKSPN
OAAITGAGTNSAIDLRQSGSKLAVAQEGSSNAKIDQTDVNLAYIDQAGSANDA
SISQAGVGTNMIQKSGNKNITQYGTQKTAIVVQSGMAIVTOR"
1253..1711
/gene="csgA"
1253..1711
/gene="csgA"
/codon_start=1
/transl_table=11
/product="CsgA protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKLLKVAIAIAIVFSSALAGVVPQYGGGAGHGGGANNSPNS
EINIYQGGNSALALQADARNSDLTQHGCGGADVGQSDSDSIDLTRQFGNSA
TLQNMGNDSHTVQKFGGNGAAYDVDTASNSIVNVQFGNNATAHQY"

ORIGIN
Query Match
Best Local Similarity 72.6%; Score 331.2; DB 1; Length 1711;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTGTTATTTATGATGTTCACAACTACGCGGTGCGCTGGATTGCAACC 60
DB 757 ATGAAAAACAATTGTTATTTATGATGTTCACAACTACGCGGTGCGCTGGATTGCAACC 60
QY 61 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
DB 817 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
QY 817 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
DB 877 CGCAACAATTATGATCTGCTGCTTCAGATATATTTTCCGTAATGATTAATGACCAAG 120
QY 121 TCTTCATTTATATCAGCGCCATTTATGTCAGAGTGGCGAGATATATGTCAGAGTA 180
DB 877 TCTTCATTTATATCAGCGCCATTTATGTCAGAGTGGCGAGATATATGTCAGAGTA 180
QY 181 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
DB 937 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
QY 937 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
DB 997 CGCAGAGAGGATATAAATCTATGTCGTTATTTTCAACAAGAGAGAAATATATCGGCG 240
QY 241 AAAGTCAGCAGGAGGAGGAAATTTATTAATCTTGGCTATATGACCAACCGGCATCCAC 300
DB 997 AAAGTCAGCAGCAGGAGGAAATTTATTAATCTTGGCTATATGACCAACCGGCATCCAC 300
QY 301 GATGCCGATATTCGCAAGAGCGCTTACGTTAATGATGACGATTTATTCAGAAAGTTCT 360
DB 1057 GATGCCGATATTCGCAAGAGCGCTTACGTTAATGATGACGATTTATTCAGAAAGTTCT 360
QY 361 GGAATTAAGCCCAATTTACCACTACGTAAGTACGAGCAAGAAACGCGTTGTGTCAGAAA 420
DB 1117 GGAATTAAGCCCAATTTACCACTACGTAAGTACGAGCAAGAAACGCGTTGTGTCAGAAA 420
QY 421 CAGTGCATATGCTTATTCGCTGCTACCCACCGCTAA 456
DB 1177 CAGTGCATATGCTTATTCGCTGCTGCTGACACACGCTAA 1212

RESULT 15
AE005315 10190 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
DEFINITION AE005315 AE005174
ACCESSION AE005315
VERSION AE005315.1 GI:12514572

```

KEYWORDS	Escherichia coli O157:H7 EDL933
SOURCE	Escherichia coli O157:H7 EDL933
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1 (bases 1 to 10190)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
JOURNAL	Nature 409 (6819), 529-533 (2001)
MEDLINE	21074935
PUBMED	11206551
REFERENCE	2 (bases 1 to 10190)
AUTHORS	Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.
TITLE	Direct Submission
JOURNAL	Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES	Location/Qualifiers
SOURCE	1..10190
gene	/organism="Escherichia coli O157:H7 EDL933"
	/mol_type="genomic DNA"
	/strain="EDL933"
	/serotype="O157:H7"
	/db_xref="taxon:155864"
	/note="enterohaemorrhagic"
	617..1072
	/gene="csgB"
	/note="synonym: Z1675"
	617..1072
	/gene="csgB"
	/function="structural component; Cell exterior constituents: Surface structures"
	/note="Residues 1 to 151 of 151 are 100.00 pct identical to residues 1 to 151 of 151 from Escherichia coli K-12 Strain MG1655: B1041"
	/codon_start=1
	/transl_table=1
	/product="minor curlin subunit precursor, similar to CsgA"
	/protein_id="AAG5787.1"
	/db_xref="GI:12514573"
	/translation="MKKKLLFMMLTIIIGAPGIAAGVGLANSEYNFAVNELSKSSFN OAATIGAGTNSAQLRGSGSKLAVVAQESSRRAKIDQDGYRLAYIDAGSANDA SISQAGATNTAMITIGKSGNANITGTQKTAIVQKOSOMAIRVOR"
	1113..1571
	/gene="csgA"
	/note="synonym: Z1676"
	1113..1571
	/gene="csgA"
	/function="regulator; Cell exterior constituents: Surface structures"
	/note="Residues 1 to 152 of 152 are 96.71 pct identical to residues 1 to 151 of 151 from Escherichia coli K-12 Strain MG1655: B1042"
	/codon_start=1
	/transl_table=1
	/product="curlin major subunit, coiled surface structures; cryptic"
	/protein_id="AAG5788.1"
	/db_xref="GI:12514574"
	/translation="MKLIKVAIAAIVPSSGALAGVPOVGGGNGGGGNSGPNSS EATIVYGGGNSALALQADANSPDITTTGGGAGADVGQSDSDSITDLTRGNSA TLDDWNGKDSHMTYKQFGGNGAAVDGTASNTVNTVQVGGNATAHQY"
	1630..1962
	/gene="csgC"
CDS	/note="synonym: Z1677"
	1630..1962
	/gene="csgC"
	/function="putative structure; Not classified"
	/note="Residues 1 to 110 of 110 are 98.18 pct identical to residues 1 to 110 of 110 from Escherichia coli K-12 Strain MG1655: B1043"
	/codon_start=1
	/transl_table=1
	/product="putative curli production protein"
	/protein_id="AAG5789.1"
	/db_xref="GI:12514575"
	/translation="MNALLALSSQITFTNTQQGDMYTIIEVTLTQSLCVOIL SIPEGSGSGQTKQEKLSLPANQPIALTSLNISISDDKVIYVTSDGSLHSQQ WPSSEKS"
	2083..2394
	/gene="Z1678"
	2083..2394
	/gene="Z1678"
	/function="orf; Unknown function"
	/note="Residues 1 to 103 of 103 are 98.05 pct identical to residues 1 to 103 of 103 from Escherichia coli K-12 Strain MG1655: B1044"
	/codon_start=1
	/transl_table=1
	/product="orf, hypothetical protein"
	/protein_id="AAG5790.1"
	/db_xref="GI:12514576"
	/translation="MFRPFLNLSLMLGSLFPPFLIAGSTAGGVYHFGQIVBPCDV STQSSPVEMNCQNGSVPGKTVSSKALMSGVNAQIASVYQYLDKOKLAVNIEY N"
	2489..3022
	/gene="Z1679"
	2489..3022
	/gene="Z1679"
	/function="orf; Unknown function"
	/note="Residues 1 to 177 of 177 are 100.00 pct identical to residues 1 to 177 of 177 from Escherichia coli K-12 Strain MG1655: B1045"
	/codon_start=1
	/transl_table=1
	/product="putative polyprotein"
	/protein_id="AAG5791.1"
	/db_xref="GI:12514577"
	/translation="MKTRIHVQGDITKLAVDVIVANANPSLKGCGVDCAIRAAP ALUDACLKVRQOQDCEPTGHAIVITLADLPAAKVHVGVMRGGEONEDQLDAYL NSRLVAVANSYTSVAPPAISTGVYPRAAAEIAVTVSEFTRHALPVOYFVCYD EENAHLYERLLTQOGDE"
	2964..4445
	/gene="ymdC"
	/note="synonym: Z1680"
	2964..4445
	/gene="ymdC"
	/function="putative enzyme; Not classified"
	/note="Residues 1 to 493 of 493 are 99.59 pct identical to residues 1 to 493 of 493 from Escherichia coli K-12 Strain MG1655: B1046"
	/codon_start=1
	/transl_table=1
	/product="putative synthase"
	/protein_id="AAG5792.1"
	/db_xref="GI:12514578"
	/translation="MKKTPSTKDSIPKKNDDPRLAAVNPICQHPGCGGLFPL EKSUDAPAAKYRLAEMSEHTLDQYITWQDMSGRLLFSALLAAKGVAVRLDDN NTPGLDITLALDLSHPRIEVRLLNPSEFLRLPLGYITDPSRLNRRMANSFTVDGVY TLVGRNITGDVAGAGEEPLSDLDVVALGPVVEDVADFPARYWCXSVPLQOQVLDV PEGEMADRIELPASMHNDAMTHRYLRMESSPFINHLVDGTLPLIMAKTKLLSDPPAK GEGKAKRHSLLPORLPIDMGSPSERIDIISSYVPYRAGVAGOLLBMVRKXVKVAILN SLAANVAVYAAAYARKKLLAYGVEIYELKTPRROSSTLHNRGITGNSGSIHART FSDIGKTVFGSFPNDRSTLNTMGFTVBSSTLQALDKRFTIQSQYDAAMQURLDR WGRINWDRRAKKEIYLKKEPATSPFKRWVVRLLSLIPVEMLL"
	4453..5610
	/gene="ymdD"

CDS
/note="synonym: Z1681"
/complement (4453..5610)
/gene="ymd"
/function="orf; Unknown function"
/note="Residues 1 to 385 of 385 are 99.22 pct identical to
residues 1 to 385 of 385 from *Escherichia coli* K-12 strain
MG1655; B1047"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55793.1"
/db_xref="GI:12514579"
/translation="MNPVADREYEXDISIRAWIMLGLIPHSILYSHTMVAAP
SLMLTLFNDPFIHSFMQVFPVIGSVSWLFLRYPLKMKRYRVRVGIPIMLTAPL
TLPOFILOYVKGKASMPGLSDYDYNLAEILSHMLFLVLVNTLLCWLAFKRI
RNLENSDKTNKFSNVKLSVFLCIGIVIRITIFLVPILISNGTFNFIWQTL
FYLPFIILALAFIFPHLKFALFTTBSRGCTLAALAFAVALNCRYSGDAMMYETS
VITWVGLMMVNVVFSFGHRLNFOARVTFVNASLFIYLVHPLTLFPGAYITPHI
TSNWLGLGLIPVVGIALIYEIHLRIPLKFLPSGKPVVRENDKAPAR"
6003..7538
/gene="mdog"
/note="synonym: Z1683"
6003..7538
/gene="mdog"
/function="enzyme; Osmotic adaptation"
/note="Residues 1 to 511 of 511 are 99.80 pct identical to
residues 1 to 511 of 511 from *Escherichia coli* K-12 strain
MG1655; B1048"
/codon_start=1
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="AAG55794.1"
/db_xref="GI:12514580"
/translation="MMKRWLSAAVNLTYSSMARSIDVAKOASLAKGVEAPK
SNLPEVPRDMKYADYOOIOPNHDKAYNNLKTPTKLEPHQCMFTDVKINEYATA
VKRIKSPDYFTFGDVQDQKDYVDLQAGKFLVLPINSKDKNDIVSLGASVFAV
GAGOVGGSAGLALDTRALPSGEFPRKFEWIERPKPTDKRLITVALDSPRATGAY
KFWVPGKDIVDVOSKIYLRDKVKGKLGVAPLTSMPLFGPQPSPPANNRPELHDSNG
LSIHAGNGEWIMRPLNPKHLAVSSFSMENPOGFLQGRDFSRFEDLDRLYDRS
AMVTPKGEWGSGVELVEIPTNDETNNIVAVTDPDLPEPGKEMNFKYITFSRDE
KLHAPDNAMVOQTRRSQDVOSNLRIPOGTLIAFVDFPGAKEMKIPEDTPYATQS
IGDNGEIVESTVRNPTVKGRVLMRYKVKDKAKTTMRALVAVADTLSETWSYDLP
ANE"
7531..10074
/gene="mdoh"
/note="synonym: Z1684"
7531..10074
/gene="mdoh"
/function="enzyme; Osmotic adaptation"

Query Match 72.6%; Score 331.2; DB 1; Length 10190;
Best Local Similarity 82.9%; Pred. No. 2.3e-86;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

gene
CDS
/note="synonym: Z1681"
/complement (4453..5610)
/gene="ymd"
/function="orf; Unknown function"
/note="Residues 1 to 385 of 385 are 99.22 pct identical to
residues 1 to 385 of 385 from *Escherichia coli* K-12 strain
MG1655; B1047"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55793.1"
/db_xref="GI:12514579"
/translation="MNPVADREYEXDISIRAWIMLGLIPHSILYSHTMVAAP
SLMLTLFNDPFIHSFMQVFPVIGSVSWLFLRYPLKMKRYRVRVGIPIMLTAPL
TLPOFILOYVKGKASMPGLSDYDYNLAEILSHMLFLVLVNTLLCWLAFKRI
RNLENSDKTNKFSNVKLSVFLCIGIVIRITIFLVPILISNGTFNFIWQTL
FYLPFIILALAFIFPHLKFALFTTBSRGCTLAALAFAVALNCRYSGDAMMYETS
VITWVGLMMVNVVFSFGHRLNFOARVTFVNASLFIYLVHPLTLFPGAYITPHI
TSNWLGLGLIPVVGIALIYEIHLRIPLKFLPSGKPVVRENDKAPAR"
6003..7538
/gene="mdog"
/note="synonym: Z1683"
6003..7538
/gene="mdog"
/function="enzyme; Osmotic adaptation"
/note="Residues 1 to 511 of 511 are 99.80 pct identical to
residues 1 to 511 of 511 from *Escherichia coli* K-12 strain
MG1655; B1048"
/codon_start=1
/transl_table=11
/product="periplasmic glucans biosynthesis protein"
/protein_id="AAG55794.1"
/db_xref="GI:12514580"
/translation="MMKRWLSAAVNLTYSSMARSIDVAKOASLAKGVEAPK
SNLPEVPRDMKYADYOOIOPNHDKAYNNLKTPTKLEPHQCMFTDVKINEYATA
VKRIKSPDYFTFGDVQDQKDYVDLQAGKFLVLPINSKDKNDIVSLGASVFAV
GAGOVGGSAGLALDTRALPSGEFPRKFEWIERPKPTDKRLITVALDSPRATGAY
KFWVPGKDIVDVOSKIYLRDKVKGKLGVAPLTSMPLFGPQPSPPANNRPELHDSNG
LSIHAGNGEWIMRPLNPKHLAVSSFSMENPOGFLQGRDFSRFEDLDRLYDRS
AMVTPKGEWGSGVELVEIPTNDETNNIVAVTDPDLPEPGKEMNFKYITFSRDE
KLHAPDNAMVOQTRRSQDVOSNLRIPOGTLIAFVDFPGAKEMKIPEDTPYATQS
IGDNGEIVESTVRNPTVKGRVLMRYKVKDKAKTTMRALVAVADTLSETWSYDLP
ANE"
7531..10074
/gene="mdoh"
/note="synonym: Z1684"
7531..10074
/gene="mdoh"
/function="enzyme; Osmotic adaptation"

Query 1 ATGAAAAACAATGTTATTATGATGATGACAAATACTGGTGGCTGGGATTGCAACC 60
Db 617 ATGAAAAACAATGTTATTATGATGATGACAAATACTGGTGGCTGGGATTGCAACC 676
Qy 61 GCGACAAATTATGATCGCTCGTTGCAATATATTTTGGGTAATATGAAATTAAGCAAG 120
Db 677 GCGACAGTTATGATTTAGCTTAATTCAGAAATATTAATCTGCGGTAATATGATTAAG 736
Qy 121 TCTTCATTATACAGGGGCGCATTTGTCAGAGTGGCAGCGATATATAGTCCAGAGTA 180
Db 737 TCTTCATTATACAGGCGCATTAATTTGTCAGAGTGGGCTAAATATAGTCTCAGTTA 796
Qy 181 GCGCAGAGAGATCAAAATCTATGTCGTTATTTCAAGAAGAGAGAAATATAGCGGCG 240
Db 797 GCGCAGGAGAGCTCAAAATCTTTGGCGTTGTTGCCAAGAAAGTATGTAACCGGCGCA 856
Qy 241 AAGTGCAGCAGCAGGGAATTAATCTTTCGATATTTAGCAAAACGCGCAATGCAAC 300
Db 857 AAGATTGACCAAGCAGAGATTAATCTTTCGATATTTAGTCAAGCGGCGAGTGCAT 916

Qy 301 GATGCAATATATGCAAGGCTTACGTTATATAGTACACTTATTCAGAAAGTTCT 360
Db 917 GATGCCAGTATTTGCAAGGCTTATGATTAATCTGCAATGATTCAGAAAGTTCT 976
Qy 361 GGAATTAAGGCAATATTTACAGTACGTTACGACAGAAACACAGATTGATGACAGAA 420
Db 977 GGTATTAAGCAATATTTACAGTATGTAATCAAGTATGTAATCAAGGCAATTTGATGACAGAA 1036
Qy 421 CAGTGCATATGCTATTTGCGGCTCACCACGCTAA 456
Db 1037 CAGTGCATATGCTATTTGCGGCTCACCACGCTAA 1072

Search completed: March 15, 2004, 22:49:43
Job time : 1967.17 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456
Sequence: 1 atgaataacaatgttattc.....ctgcgcaccacgaactaa 456

Scoring table: IDENTITY_NUC
Gap 10.0, Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database: N_Geneseq_29Jan04:*

- 1: Geneseq1980s:*
- 2: Geneseq1990s:*
- 3: Geneseq2000s:*
- 4: Geneseq2001as:*
- 5: Geneseq2001bs:*
- 6: Geneseq2002s:*
- 7: Geneseq2003as:*
- 8: Geneseq2003bs:*
- 9: Geneseq2003cs:*
- 10: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	3 AAC64618	AAC64618 Salmonella
2	332.8	73.0	456	3 AAC64620	AAC64620 Escherichia
3	74.4	16.3	100	7 ACD68810	ACD68810 E. coli K
4	66.4	14.6	100	7 ACD68812	ACD68812 E. coli K
5	64.8	14.2	100	7 ACD68811	ACD68811 E. coli K
6	37.6	8.2	456	3 AAC64623	AAC64623 Agfa::PT3
7	37.6	8.2	2000	7 ADA71938	ADA71938 Rice gene
8	35.8	7.9	39	3 AAC64608	AAC64608 Agfa (SEF)
9	35	7.7	738	6 AAD32400	AA32400 Chlamydia
10	35	7.7	105325	6 ABR94407	ABR94407 DNA encod
11	34.4	7.5	2006	9 ADC10081	ADC10081 Human NOV
12	34.4	7.5	18540	8 ACD26496	ACD26496 Human tum
13	34.2	7.5	456	3 AAC64629	AAC64629 Agfa::PT3
14	34	7.5	361	2 AAQ73066	AAQ73066 Agfa sequ
15	34	7.5	361	2 AAT74141	AAT74141 Salmonella
16	34	7.5	456	2 AAQ87467	AAQ87467 Agfa sequ
17	34	7.5	456	2 AAT74142	AAT74142 Salmonella
18	34	7.5	456	3 AAC64628	AAC64628 Agfa::PT3
19	34	7.5	456	3 AAC64622	AAC64622 Agfa::PT3
20	34	7.5	456	3 AAC64617	AAC64617 Salmonella
21	34	7.5	456	3 AAC64625	AAC64625 Agfa::PT3
22	34	7.5	456	3 AAC64624	AAC64624 Agfa::PT3
23	34	7.5	2000	7 ADA71938	ADA71938 Rice gene

C	24	33.8	7.4	203654	7 ABX16034	ABX16034 Human gen
C	25	33.4	7.3	53585	2 AAX20251	AAX20251 Borrelia
C	26	33.2	7.3	535	2 ADD29528	ADD29528 Mouse cae
C	27	33	7.2	33	3 AAC64614	AAC64614 S. enteri
C	28	33	7.2	3699	6 ABN70306	ABN70306 Streptococ
C	29	33	7.2	3699	6 ABN70834	ABN70834 Streptococ
C	30	33	7.2	3717	6 ABN66453	ABN66453 Streptococ
C	31	33	7.2	110000	6 ABN71527_03	Continuation (4 of
C	32	32.8	7.2	418	3 AAH30547	AAH30547 Human col
C	33	32.8	7.2	1083	9 ADC10077	ADC10077 Human NOV
C	34	32.8	7.2	1098	9 ADC10079	ADC10079 Human NOV
C	35	32.8	7.2	1105	9 ADC10075	ADC10075 Human NOV
C	36	32.8	7.2	1168	9 ADC10069	ADC10069 Human NOV
C	37	32.8	7.2	1171	9 ADC10071	ADC10071 Human NOV
C	38	32.8	7.2	1299	9 ADC10073	ADC10073 Human NOV
C	39	32.8	7.2	1938	9 ADC10067	ADC10067 Human NOV
C	40	32.8	7.2	2649	7 ADA53932	ADA53932 Human cod
C	41	32.8	7.2	2843	7 ACC44405	ACC44405 Gene enco
C	42	32.4	7.1	779	1 AAN40272	AAN40272 Sequence
C	43	32.4	7.1	1914	4 AAS51585	AAS51585 Staphyloc
C	44	32.4	7.1	1929	4 AAS54408	AAS54408 Staphyloc
C	45	32.4	7.1	10732	3 AAA10594	AAA10594 Gene enco

ALIGNMENTS

RESULT 1
ID AAC64618 standard; DNA; 456 BP.
XX
AC AAC64618;
XX
DT 26-FEB-2001 (first entry)
XX
DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:2.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
PN WO200060102-A2.
XX
FD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
(UTVI-) UNIV VICTORIA.
XX
White AP, Doran JL, Collison SK, Kay MW;
XX
WPI: 2000-672631/65.
XX
P-PSDB; AAB36342.
XX
Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
XX
The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 8,6e-135;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATGATGTTGCAATCTGCGTGGCGATTTGCAACC 60
 DB 1 ATGAAAAACAATTTGTTATGATGTTGCAATCTGCGTGGCGATTTGCAACC 60
 QY 61 GCGCAAAATTTGATCTGCGTCTGCTGCAATATATTTGCGGTAATTAAGCAAG 120
 DB 61 GCGCAAAATTTGATCTGCGTCTGCTGCAATATATTTGCGGTAATTAAGCAAG 120
 QY 121 TCTTCATTATATGCGGGGCGCATTTGTTGCAAGTGGCGAGATTAAGTCCAGAGTA 180
 DB 121 TCTTCATTATATGCGGGGCGCATTTGTTGCAAGTGGCGAGATTAAGTCCAGAGTA 180
 QY 181 CGCGAGAGAGATCAAAATATTTGCTGTTATTTGCAAGAGAGGAAATTAATCGGCGC 240
 DB 181 CGCGAGAGAGATCAAAATATTTGCTGTTATTTGCAAGAGAGGAAATTAATCGGCGC 240
 QY 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGCTATTTGAGCAAAACCGGCAATGCCAAC 300
 DB 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGCTATTTGAGCAAAACCGGCAATGCCAAC 300
 QY 301 GATGCCAGATATTCGCAAGCGCTTACGTAATAGTGAAGTATTAATCCAGAAAGTTCT 360
 DB 301 GATGCCAGATATTCGCAAGCGCTTACGTAATAGTGAAGTATTAATCCAGAAAGTTCT 360
 QY 361 GGAATTAAGGCAATATTAACCAAGCGTACGCAAGAAACAGCAAGTTAGTGCAGAA 420
 DB 361 GGAATTAAGGCAATATTAACCAAGCGTACGCAAGAAACAGCAAGTTAGTGCAGAA 420
 QY 421 CAGTGGCATATGCGTATTCGCGTACCCAGCGCTAA 456
 DB 421 CAGTGGCATATGCGTATTCGCGTACCCAGCGCTAA 456

RESULT 2
 AAC64620
 ID AAC64620 standard; DNA; 456 BP.
 AC AAC64620;
 XX 26-FEB-2001 (first entry)
 DE *Escherichia coli* CsgB DNA sequence SEQ ID NO:4.
 XX *Escherichia coli* CsgB DNA sequence SEQ ID NO:4.
 KM *Salmonella*; *agfa*; Chromosomal gene replacement; fibrin; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Escherichia coli*.
 XX MO200060102-A2.
 PN
 XX

PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUYV-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay MW;
 DR WPI: 2000-672631/65.
 XX P-BSDb; AAB36344.
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 134; 139p; English.

The present invention describes a recombinant agfa gene (1) where a
 segment of the gene has been replaced by a segment of a foreign DNA
 sequence which encodes a foreign epitope or antigen. Also described are:
 (1) use of thin aggregative fibrillae (SEF17/TAFF) nucleation depended
 assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fibrillae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fibrillar presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fibrin protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fibrillar subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fibrillae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;

Query Match 73.0%; Score 332.8; DB 3; Length 456;
 Best Local Similarity 83.1%; Pred. No. 1.5e-95;
 Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATGATGTTGCAATCTGCGTGGCGATTTGCAACC 60
 DB 1 ATGAAAAACAATTTGTTATGATGTTGCAATCTGCGTGGCGATTTGCAACC 60
 QY 61 GCGCAAAATTTGATCTGCGTCTGCTGCAATATATTTGCGGTAATTAAGCAAG 120
 DB 61 GCGCAAAATTTGATCTGCGTCTGCTGCAATATATTTGCGGTAATTAAGCAAG 120
 QY 121 TCTTCATTATATGCGGGGCGCATTTGTTGCAAGTGGCGAGATTAAGTCCAGAGTA 180
 DB 121 TCTTCATTATATGCGGGGCGCATTTGTTGCAAGTGGCGAGATTAAGTCCAGAGTA 180
 QY 181 CGCGAGAGAGATCAAAATATTTGCTGTTATTTGCAAGAGAGGAAATTAATCGGCGC 240
 DB 181 CGCGAGAGAGATCAAAATATTTGCTGTTATTTGCAAGAGAGGAAATTAATCGGCGC 240
 QY 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGCTATTTGAGCAAAACCGGCAATGCCAAC 300
 DB 241 AAAGTCAGCAGCAGCAGGAAATTAATCTTGGCTATTTGAGCAAAACCGGCAATGCCAAC 300
 QY 301 GATGCCAGATATTCGCAAGCGCTTACGTAATAGTGAAGTATTAATCCAGAAAGTTCT 360

```
Db 301 GATGCCATATTTGCGAAGTGTCTTATGTAATATGCGATGATTTATCCAGAAAGTTCT 360
Oy 361 GGAATAAAGCCCAATATTATCCAGTACGCGAAGCAAGCAAGTTGATGTCAGAAA 420
Db 361 GGTATATAAAGCAATATTATACAGATAGTGTACTCAAAAACGGCAATGTTAGTGCAGAGA 420
Oy 421 CAGTCGCAATGCGTATGCGGTGACCCCAAGCTTAA 456
Db 421 CAGTCGCAATGCGTATGCGGTGACACACAGTTAA 456

RESULT 3
ACD68810
ID ACD68810 standard; DNA; 100 BP.
XX
AC ACD68810;
XX
XX 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
XX Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
XX EPI260592-A1.
XX
XX 27-NOV-2002.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX (MWGB-) MWG-BIOTECH AG.
XX
XX Donner H, Drescher B, Huber A, Weber J;
XX
XX WPI: 2003-241155/24.
XX
XX Biochip containing probes complementary with open reading frames in
XX Escherichia coli K12, useful for detecting gene expression and expression
XX patterns.
XX
XX Claim 3; Page 23; 2004pp; German.
XX
XX This invention describes a novel biochip comprising probe spots, each
XX containing many identical probes. The probes are nucleotide sequences of
XX 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
XX least one includes a segment of at least 20 bases identical with, or
XX complementary to, a segment of an open reading frame (orf) of Escherichia
XX coli K12. The biochip is used for specific detection of gene expression
XX in K12 and for determining the gene expression pattern, e.g. for
XX diagnostic determination of which E. coli strains are present in the gut,
XX and to determine the effects of e.g. growth media on gene expression. The
XX biochip provides as comprehensive as possible detection of the K12
XX genome, with simultaneous analysis of many different genes with a single
XX device, and comparison of gene expression between K12 and its mutants or
XX other E. coli strains in a single experiment. Apart from qualitative and
XX quantitative information about gene expression, it also allows
XX measurements of population about gene expression. The use of
XX synthetic oligonucleotides for preparation of probes allows free
XX variation in probe length and ensures high purity (and thus selectivity,
XX reactivity and reproducibility); also synthetic probes are generally
XX shorter than probes prepared by polymerase chain reaction. ACD68731 to
XX ACD81540 represent oligonucleotide probes used with the biochip described
XX in the invention
XX
XX Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;
XX
Query Match 16.3%; Score 74.4; DB 7; Length 100;
Best Local Similarity 84.0%; Pred.No. 1.6e-13;
Matches 84; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
```

```
Oy 266 ACTTGGCATATTTGAGCAAAAGGCGCATGCGACGATCGATATATCGCAAGCGTT 325
Db 1 ACCTTGATATATTGATCAGCGCGGCGAGTGCACGATGCGATTTTCGCAAGTGTCT 60
Oy 326 ACGGTAATAGTGCAGCTATTATTCAGAAAGGTTCTGAGAA 365
Db 61 ATGGTAATACTGCGATGATTTATTCAGAAAGGTTCTGGTAA 100

RESULT 4
ACD68812
ID ACD68812 standard; DNA; 100 BP.
XX
AC ACD68812;
XX
XX 18-SEP-2003 (first entry)
XX
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
XX
XX Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
XX
XX Escherichia coli.
XX
XX EPI260592-A1.
XX
XX 27-NOV-2002.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX 17-MAY-2001; 2001EP-00112179.
XX
XX (MWGB-) MWG-BIOTECH AG.
XX
XX Donner H, Drescher B, Huber A, Weber J;
XX
XX WPI: 2003-241155/24.
XX
XX Biochip containing probes complementary with open reading frames in
XX Escherichia coli K12, useful for detecting gene expression and expression
XX patterns.
XX
XX Claim 3; Page 23; 2004pp; German.
XX
XX This invention describes a novel biochip comprising probe spots, each
XX containing many identical probes. The probes are nucleotide sequences of
XX 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
XX least one includes a segment of at least 20 bases identical with, or
XX complementary to, a segment of an open reading frame (orf) of Escherichia
XX coli K12. The biochip is used for specific detection of gene expression
XX in K12 and for determining the gene expression pattern, e.g. for
XX diagnostic determination of which E. coli strains are present in the gut,
XX and to determine the effects of e.g. growth media on gene expression. The
XX biochip provides as comprehensive as possible detection of the K12
XX genome, with simultaneous analysis of many different genes with a single
XX device, and comparison of gene expression between K12 and its mutants or
XX other E. coli strains in a single experiment. Apart from qualitative and
XX quantitative information about gene expression, it also allows
XX measurements of population about gene expression. The use of
XX synthetic oligonucleotides for preparation of probes allows free
XX variation in probe length and ensures high purity (and thus selectivity,
XX reactivity and reproducibility); also synthetic probes are generally
XX shorter than probes prepared by polymerase chain reaction. ACD68731 to
XX ACD81540 represent oligonucleotide probes used with the biochip described
XX in the invention
XX
XX Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;
XX
Query Match 14.6%; Score 66.4; DB 7; Length 100;
Best Local Similarity 79.0%; Pred.No. 5.6e-11;
Matches 79; Conservative 0; Mismatches 21; Indels 0; Gaps 0;
```

Db 1 TTGATTTAATCAGCGACCCATTAATGTCAGTGGACTAATATAGTGTGCTAGTACG 60
OY 183 CCGAGGAGATCAAACTATGTCCTGTTATTTACAGAGA 222
Db 61 GCAGGAGGCTCAAACTTTGGCGGTGTGTCGAGAGA 100

RESULT 5
ACD68811
ID ACD68811 standard; DNA; 100 BP.
XX
AC ACD68811;
XX
DT 18-SEP-2003 (first entry)
XX
DE E. coli K12 MG1655 biochip probe SEQ ID 81.
XX
KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS Escherichia coli.
XX
PN EP1260592-A1.
XX
PD 27-NOV-2002.
XX
PF 17-MAY-2001; 2001EP-00112179.
XX
PR 17-MAY-2001; 2001EP-00112179.
XX
PA (MWGB-) MWG-BIOTECH AG.
XX
PI Donner H, Drescher B, Huber A, Weber J;
XX
DR WPI; 2003-241155/24.
XX
PT Biochip containing probes complementary with open reading frames in
PT Escherichia coli K12, useful for detecting gene expression and expression
XX patterns.

Claim 3; Page 23; 2004pp; German.

CC This invention describes a novel biochip comprising probe spots, each
CC containing many identical probes. The probes are nucleotide sequences of
CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC least one includes a segment of at least 20 bases identical with, or
CC complementary to, a segment of an open reading frame (orf) of Escherichia
CC coli K12. The biochip is used for specific detection of gene expression
CC in K12 and for determining the gene expression pattern, e.g. for
CC diagnostic determination of which E. coli strains are present in the gut,
CC and to determine the effects of e.g. growth media on gene expression. The
CC biochip provides as comprehensive as possible detection of the K12
CC genome, with simultaneous analysis of many different genes with a single
CC device, and comparison of gene expression between K12 and its mutants or
CC other E. coli strains in a single experiment. Apart from qualitative and
CC quantitative information about gene expression, it also allows
CC measurements of population densities for the various strains. The use of
CC synthetic oligonucleotides for preparation of probes allows free
CC variation in probe length and ensures high purity (and thus selectivity,
CC reactivity and reproducibility); also synthetic probes are generally
CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC ACD81540 represent oligonucleotide probes used with the biochip described
CC in the invention

XX
XX
SQ Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

Query Match 14.2%; Score 64.8; DB 7; Length 100;
Best Local Similarity 78.0%; Pred. No. 1.8e-10;
Matches 78; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

OY 208 GTTATTTCACAGAGAGAGAAATATCGGGGAGAAAGTCGACCGAGGAGGAATTATTAAC 267
Db 1 GTTGTGGCCAGAGAGATGATGCAACCGGGCAAGATTGACGACAGAGAGATTATTAAC 60

OY 268 TTGGCTATATTTGAGCAACGGGCAATGCCAATGCCAATGCCA 307
Db 61 CTTGCATATATTGATCAGCGCGGCAATGCCAATGCCAATGCCA 100

RESULT 6
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 DNA sequence SEQ ID NO:13.
XX
KM Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
PN W0200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNITV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR WPI; 2000-672631/65.
XX
PT P-PSDB; AAB36347.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.

PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP17/7AP) nucleation depended
CC assembly system of strains of Salmonella (Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

	Query Match	8.2%;	Score 37.6;	DB 3;	Length 456;	
	Best Local Similarity	51.2%;	Pred. No. 0.15;	Mismatches 84;	Indels 0;	Gaps 0
	Matches	88;	Conservative	0;		
Oy	272	CGTATATTGACGAAGCGGCATGCGCAATGCGCATGTATATCGCAAAACGGCTTAGCGTA	331			
Dd	203	CGACCATTATCCAGAGCGGTTATGTGTAACGGGCCGCGATATAGCCAGSGTGCGATATA	262			
Oy	332	ATAGTCAGCTATTATTCAGAAAAGTCTCGAAATPAGGCCAATATTACCCAGTACGGTA	391			
Dd	263	GTACTATTGAAGTGAAGTCAAGATGGTTTCAGAAATATATSCACACATCGACAGTGGAACG	322			
Oy	392	CGCAGAAAACAGCACTGTGTAGTCAGAAAAACATCGCATATGGCTATTGGCGGT	443			
Dd	323	CTAAAACTCCGATATATCTGTGCGCCCAATATGATCACTGGTACCCGTGT	374			
RESULT 7						
ID	ADA71938/C					
AD	ADA71938 standard; DNA; 2000 BP.					
XX	ADA71938;					
XX	20-NOV-2003 (first entry)					
DE	Rice gene, SEQ ID 5263.					
KW	Plant; bacterial infection; fungal infection; viral infection; rice;					
XX	gene; ds.					
OS	Oryza sativa.					
XX	WO200300898-A1.					
PN	03-JAN-2003.					
PD	22-JUN-2001; 2001WO-IB001105.					
PP	22-JUN-2001; 2001WO-IB001105.					
XX	22-JUN-2001; 2001WO-IB001105.					
PR	(SYGN) SYNGENTA PARTICIPATIONS AG.					
PA	Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,					
PI	Katagiri F, Quan S, Tao Y, Whitlam S, Xie Z, Zhu T, Zou G;					
XX	WP; 2003-175290/17.					
DR						
PT	Identifying at least one gene involved in plant resistance or response to					
PT	pathogenic infection for conferring resistance or tolerance to a plant to					
PT	bacterial, fungal or viral infection by determining or detecting plant					
PT	gene expression.					
XX						
PS	Claim 27; SEQ ID NO 5263; 899pd; English.					
XX						
CC	The present invention relates to a method (M1) for identifying genes					
CC	involved in plant resistance or response to pathogenic infection. M1					
CC	comprises identifying a gene whose expression is significantly altered in					
CC	the incompatible interaction of plant gene expression relative to					
CC	expression of the gene in an uninfected plant, in a mutant plant that					
CC	does not express a gene associated with response to pathogenic infection,					
CC	or in a corresponding incompatible or compatible interaction. (M1) is					
CC	useful for conferring resistance to resistance or tolerance to a plant to					
CC	bacterial, fungal or viral infection. The present sequence was used to					
CC	illustrate the invention.					
SQ	Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;					
Query Match	8.2%; Score 37.6; DB 7; Length 2000;					
Best Local Similarity	12.8%; Pred. No. 0.26;					
Matches	57; Conservative 195; Mismatches 189; Indels 3; Gaps 2					
Oy	14 TGTTATTATGATCTTGACAATACCTGCGGCGCCCTGGAGTTGCAACCGGACAAATTATG	73				

Db	1027	IGTAGTCTCTKMRIRYTSMSFTYAMKKYTKYMTAYSSTWKMYAYKRAYAMSSRKT	968
Qy	74	ATCTGGCTCGTTACAGATATAATTTTCGGCTAAAGAAATTAAGCAAGCTTCATTATATC	133
Db	967	WMCTGCKRMATYCGTKMAAGRMWRMAWCMYCCMMKMKMTSCKMKYWRWTSWYTW	908
Qy	134	AGCGGCCATTATTGTCAGTCGGCAGCATTAATATGTCGACAGTAAGCCAGAAAGAT	195
Db	907	WGANRYAYANRRRMTYKMSWRBYWMTTKAAMWMTCMCAKMYATGMAWMMWRYT	848
Qy	194	CAAACTATTGTCGTTATTTTCACAAAGAGGAAATTAATGCGGGAATGACACAG	255
Db	847	MYTCYATNCKACKCKYKAMMTKMTTWTACAMRITSRMRMAAGMRKRYMKRAYMRWR	788
Qy	254	CAGGAATTAATACTTTCGCTATATTGAGCAACGGGCAATGCCACGATGCCAGTATAT	313
Db	787	CMKAGMAMKMSRYMKMKKATATRYWKMMAATMMWSMRMHSYRWMSGGRWMSMR	728
Qy	314	CGCAAGCGCTTACGCTAATATGCACTATTATCCAGAAAGTTTGGAAATAGGCCA	373
Db	727	CS--RMKCAKTKYASSARWTKAKRSYRRRWYMKRMKMTYRRYRWSCRMTRAMSKR	670
Qy	374	ATATTACCGACGATCGTACGCA-GAAACAGACGATTGTAGTCAGAAACAGTGCATATG	432
Db	669	KKMAGSMKSCMWTWRGASRMWYSKYSCSAKCCCKTRWTSYSTGTGMTSYKSMW	610
Qy	433	GCTATTCCGCTCACCCCAACGCTAA	456
Db	609	TSKMSYMKMTCTMYTSMKGSRR	586
RESULT 8			
AAAC64608			
ID	AAAC64608	standard; DNA; 39 BP.	
XX	AAAC64608;		
AC	06-AUG-2003	(revised)	
DT	26-FEB-2001	(first entry)	
DT			
XX	Agfa (SEF17) recombinant agfa::P73a generating PCR primer 17-A.		
XX	Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;		
KW	vaccine; immune response; immunogen; PCR primer; ss.		
RW	Salmonella enteritidis.		
XX	Synthetic.		
OS			
XX	W0200060102-A2.		
FN			
XX	12-OCT-2000.		
PD			
XX	05-APR-2000; 2000WO-CA000356.		
PF	05-APR-1999; 99US-0127888P.		
XX	(UYVI-) UNIV VICTORIA.		
PA			
PI	White AP, Doran JL, Collison SK, Kay W;		
DR	WPI; 2000-672631/65.		
XX			
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence		
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa		
PT	protein useful for eliciting immune response in animal.		
XX			
XX	Example 2; Page 61; 139pp; English.		
CC	The present invention describes a recombinant agfa gene (1) where a		
CC	segment of the gene has been replaced by a segment of a foreign DNA		
CC	sequence which encodes a foreign epitope or antigen. Also described are:		
CC	(1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended		
CC	assembly system of strains of Salmonella, Escherichia coli and		

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
 CC correct OS field.)

SQ Sequence 39 BP; 13 A; 7 C; 10 G; 9 T; 0 U; 0 Other;

Query Match
 Best Local Similarity 7.9%; Score 35.8; DB 3; Length 39;
 Matches 37; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 393 GCAGAAAAACAGCTGTAGTCAGAGAACAGTGCCTAT 431
 Db 1 GCAGATTTCAGCACTTGTAGTCAGAGAACAGTGCCTAT 39

RESULT 9
 AAD32400
 ID AAD32400 standard; DNA; 738 BP.

AC AAD32400;
 DT 29-AUG-2003 (revised)
 DT 18-JUN-2002 (first entry)

DE Chlamydia pneumoniae nlpd gene.

XX Chlamydiaceae family; chronic infection; persistent infection; pyk; nlpd;
 KW Cpn0585; regulatory pathway; biosynthetic pathway; ompA; ompB; hsp60;
 KW lipopolysaccharide; cardiovascular system; respiratory tract; therapy;
 KW genital tract; reproductive system; atherosclerotic tissue; macrophage;
 KW multiple sclerosis; conjunctiva; prophylaxis; antibacterial; gene; ds.

OS Chlamydia pneumoniae.

XX Key Location/Qualifiers
 FH CDS 1..738
 FT /*tag= a
 FT /product= "Chlamydia pneumoniae nlpd protein"
 FT /note= "CDS does not include start codon"
 FT /partial

XX MO200214516-A1.

XX 21-FEB-2002.

XX 17-AUG-2001; 2001MO-AU001021.

XX 18-AUG-2000; 2000AU-00009540.

XX (YOU-) UNIV QUEENSLAND TECHNOLOGY.

XX (MATH/) MATHEWS S A.

XX Timms P;
 XX

DR MPI; 2002-269197/31.
 DR P-PSDB; AAE20289.

XX
 PT Detecting Chlamydial organism in its persistent phase by detecting
 PT expression change of range of genes belonging to their respective
 PT biosynthetic pathways when expression is compared to that of organism in
 PT lytic phase.

XX Disclosure; Page 117-118; 196pp; English.

XX The invention relates to composition and methods for detecting organisms
 CC of the Chlamydiaceae family, including species of Chlamydia and
 CC Chlamydia, in the persistent phase of their developmental cycle and for
 CC the diagnosis of chronic or persistent infections caused by such
 CC organisms. The composition is useful for modulating the expression of
 CC gene such as pyk, nlpd, Cpn0585, a gene belonging to same regulatory/
 CC biosynthetic pathway and ompA, ompB, hsp60, a gene involved in
 CC lipopolysaccharide biosynthesis. It is also useful for modulating the
 CC level and/or functional activity of an expression product of these genes,
 CC where the gene is present in an epithelial cell (selected from
 CC cardiovascular system, respiratory tract, genital tract, reproductive
 CC system or conjunctiva), macrophage, or a cell associated with
 CC atherosclerotic tissue or associated with multiple sclerosis brain
 CC tissue. The composition is useful for treatment and/or prophylaxis of a
 CC chronic infection caused by an organism of the Chlamydiaceae family in a
 CC patient. Antigen associated with the persistent phase of the
 CC in the manufacture of a medicament, for treating and/or preventing
 CC Chlamydiaceae infection in a patient. The present sequence is Chlamydia
 CC pneumoniae nlpd gene. (Updated on 29-AUG-2003 to standardise OS field)

SQ Sequence 738 BP; 231 A; 158 C; 166 G; 183 T; 0 U; 0 Other;

Query Match
 Best Local Similarity 7.7%; Score 35; DB 6; Length 738;
 Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

OY 340 GCTATTATCCAGAAAGTTCTGGAATATAGGCCATATTTCCAGTAGCGTACGAGAA 399
 Db 1 GCTTTGATCAGAGAAATGACGGCAGTGAATATGATGATGAGACATGTATATA 60

OY 400 ACAGAGTGTAGTCAGAGAAACAGTGCATATGCTATTCGCTCAC 446
 Db 61 ACAGCTTCGTAGTGAATGCTATATGCTTGTGCTTTGTGTAC 107

RESULT 10

ABK94407
 ID ABK94407 standard; DNA; 105325 BP.

AC ABK94407;

DT 27-AUG-2002 (first entry)

DE DNA encoding endothelin-3 (EDN-3).

XX Endothelin 3; EDN-3; endothelin converting enzyme; ECE;
 KW endothelin receptor; EDNR; signaling system; cardiovascular disease;
 KW coronary heart disease; hypertension; atherosclerosis; anglogenesis;
 KW fatty acid metabolism; diabetes; familial hypercholesterolemia;
 KW forensic marker; transgenic animal; solid support; SNP;
 KW cardiovascular regulator; gene; ds; single nucleotide polymorphism.

XX Homo sapiens.

XX Key Location/Qualifiers
 FH variation
 FT /*tag= a
 FT /replace_name= "Single nucleotide polymorphism"
 FT /replace(59430,C)
 FT /*tag= b
 FT /standard_name= "Single nucleotide polymorphism"
 FT replace(63843,T)

XX The invention relates to novel isolated polypeptides, mature form of the
 CC polypeptide, a sequence that is 95% identical to the polypeptide or the
 CC polypeptide comprising one or more conservative substitutions. The NOVX
 CC polypeptide is useful for treating or preventing a pathology associated
 CC with the polypeptide e.g. disorders associated with aberrant expression
 CC or activity of the polypeptide, such as cancer, diabetes, obesity, and
 CC endocrine, CNS and inflammatory disorders. They can also be used in
 CC various detection and screening assays, chromosome mapping, tissue typing
 CC and predictive medicine. This sequence corresponds to the coding sequence
 CC for one of the polypeptides of the invention.

SO Sequence 2006 BP; 637 A; 390 C; 493 G; 486 T; 0 U; 0 Other;

Query Match 7.5%; Score 34.4; DB 9; Length 2006;
 Best Local Similarity 48.9%; Pred. No. 2.7;

Matches 92; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 208 GTTATTTCACAGAGAGAAATATTCGGCGAAGTCGACGAGGAGAAATTATAC 267
 Db 540 GTTCTGATCTAGAGAGAGATATCAAAAGCTATATACGTCGACGAGGCTCTAC 599
 QY 268 TTTGGCTATTGAGCAAGGAGCAATGCCAGATCCGATATTCGCAAGCGCTTAC 327
 Db 600 CTGGTGTCTGAGAGAGCTCTGTTCAATACCTCGAAGACATAGTCAAGTATCACTA 659
 QY 328 GGTATATGTCAGCTATTATTCAGAAAGGTTCTGAAATATAGGCCAATTTACCACTAC 387
 Db 660 AATCTGCGCCCTGGGAAATATCCAGAAATATCCAGAGAAATATCACTTACCAAGGAC 719
 QY 388 GGTACGCA 395
 Db 720 TTAACGCA 727

RESULT 12

ACD26496/C

ID ACD26496 standard; DNA; 18540 BP.

AC ACD26496;

DT 09-SEP-2003 (first entry)

DE Human tumour necrosis factor receptor related protein-1, TRP-1, DNA.

XX Human; ds; gene; TRP-1; inflammatory disorder; rheumatoid arthritis;
 KW tumour necrosis factor receptor related protein-1; Crohn's disease; AIDS;
 KW asthma; irritable bowel syndrome; ulcerative colitis; multiple sclerosis;
 KW diabetes mellitus; immune disorder; autoimmune disease; allergy; stroke;
 KW systemic lupus erythematosus; neurodegenerative disorder; prostatitis;
 KW Alzheimer's disease; Parkinson's disease; cell proliferative disorder;
 KW cell differentiation disorder; ischaemic injury; myocardial infarction;
 KW apoptotic disorder; gastrointestinal and reproductive tract disorder;
 KW cancer; ulcer; bone disorder; osteoporosis; blood disorder; hepatitis;
 KW aplastic anaemia; thrombocytopenia; viral disorder; viral encephalitis;
 KW chromosome Xq12.

OS Homo sapiens.

PN US2003059889-A1.

PD 27-MAR-2003.

PF 28-AUG-2002; 2002US-00231426.

PR 05-SEP-2001; 2001US-0317151P.

PA (BODI/) BODIAN D L.

PA (LABO/) LABOW M A.

PI (MICK/) MICKANIN C S.

XX Bodian DL, Labow MA, Mickanin CS;

DR WPI, 2003-521912/49.
 DX P-PSDB; AB062739.

PT New tumor necrosis factor receptor related protein-1 polypeptide and
 PT nucleic acid, useful for diagnosing and treating disorders in a human,
 PT e.g. an inflammatory disorder, an immune disorder or a neurodegenerative
 PS disorder.
 XS Claim 16; Page 21-30; 36pp; English.

CC The invention relates to a new isolated tumour necrosis factor receptor
 CC related protein-1 polypeptide, TRP-1. The tumour necrosis factor receptor
 CC related protein-1 polypeptide, nucleic acid and antibody are useful for
 CC diagnosing and treating disorders in a human, such as an inflammatory
 CC disorder (e.g. rheumatoid arthritis, Crohn's disease, AIDS, allergies,
 CC asthma, irritable bowel syndrome, ulcerative colitis, multiple sclerosis
 CC or diabetes mellitus), an immune disorder (e.g. autoimmune disease or
 CC systemic lupus erythematosus), a neurodegenerative disorder (e.g.
 CC Alzheimer's disease or Parkinson's disease), a cell proliferative
 CC disorder (e.g. various types of cancer), a cell differentiation disorder
 CC (e.g. stroke, ischaemic injury or myocardial infarction), an apoptotic
 CC or proliferative disorder, a gastrointestinal and reproductive tract disorder (e.g. ulcers
 CC or prostaticitis), bone disorder (e.g. osteoporosis), blood disorder (e.g.
 CC aplastic anaemia or thrombocytopenia) or viral disorder (e.g. hepatitis
 CC or viral encephalitis). The polypeptides are also useful in manufacturing
 CC a medicament for treating the disorders. The methods are useful for
 CC identifying molecules that bind to and/or modulate the activity of the
 CC polypeptide, or the transcription or translation of the nucleic acid. The
 CC present sequence represents human tumour necrosis factor receptor related
 CC protein-1, TRP-1, DNA which is located on chromosome Xq12

SO Sequence 18540 BP; 5173 A; 3532 C; 3591 G; 6244 T; 0 U; 0 Other;

Query Match 7.5%; Score 34.4; DB 8; Length 18540;
 Best Local Similarity 52.0%; Pred. No. 6.5;

Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 86 CAGATATTAATTTGCGGTAATGAATTAAGCAAGTCTTCAATTAATCAGGCGCCATTA 145
 Db 5513 CAGATGTAATTAATTAATAAATGAACAGAAATTTTAATCACTGATGGGCTTA 5454
 QY 146 TTGTCAGTGGCAGCGGATATAGTCCAGATGAGCCGAGAGATCAAACTATGT 205
 Db 5453 GTAGTAGAGAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 5394
 QY 206 CCGTATTTCACAGAGAGAGCAAAATTA 233
 Db 5393 TGCATATGAAACACAGGAGAGAAATGA 5366

RESULT 13

AAC64629

ID AAC64629 standard; DNA; 456 BP.

AC AAC64629;

DT 26-FEB-2001 (first entry)

DE AGFA::PT3#8 DNA sequence SEQ ID NO:25.

XX Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
 KW vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.
XX (UVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay WW;
PI WPI: 2000-672631/65.
XX P-PSDB; AAB36353.
DR
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure: Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
CC
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
Query Match 7.5%; Score 34.2; DB 3; Length 456;
Best Local Similarity 56.8%; Pred. No. 1.8;
Matches 63; Conservative 0; Mismatches 48; Indels 0; Gaps 0;
OY 275 ATATTGACCAAGCGGCAATGCCAGATATGCGAAGCGCTTACGGTAAATA 334
DB 206 AAATGGCATGCGAGTATGTATGTAACGGCGCGATGTAGGCCAGGTCCGATTAATA 265
OY 335 GTGACGCTATTATCCAGAAAGTTCTGAAATTAAGCCAAATATATACCAGT 385
DB 266 CTATTGAACCTACAGATGTGTTTCAGAAATTAATGACCATGACACAGT 316
RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
XX AAQ73066;
XX
XX 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX *Salmonella*; Agfa; vaccine; genetic immunization; ds.
XX
XX *Salmonella enteritidis*.
XX

PH Key Location/Qualifiers
FT CDS 1..359
FT /*tag= a
FT /note= "Agfa"
FT misc_feature 37..60
FT /*tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT 52..69
FT /*tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT misc_feature complement(103..129)
FT /*tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT complement(129..402)
FT /*tag= c
FT /note= "TAF4 primer (pair with TAF3)"
FT
XX W09425598-A2.
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94WO-IB000207.
XX PF 26-APR-1994;
XX PR 26-APR-1993; 93US-00054452.
XX
XX (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL;
PI WPI: 1994-358275/44.
XX P-PSDB; AAR62761.
XX
XX
XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
XX strains, vector constructs, or compens. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The DNA encodes the *Salmonella enteritidis*27655-3b *TnphoA* mutant strain
CC agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to *Salmonella* in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
Query Match 7.5%; Score 34; DB 2; Length 361;
Best Local Similarity 56.1%; Pred. No. 1.9;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
OY 272 CGTATATTGACCAAGCGGCAATGCCAGATATGCGAAGCGCTTACGGTAA 331
DB 140 CGACCATTAACCAAGCGGCTATGTATGTAACGGCGCGATGTAGGCCAGGTCCGATTAATA 199
OY 332 ATATGACGCTATTATCCAGAAAGTTCTGAAATTAAGCCAAATATATACCAGT 385
DB 200 GTACTATTGAACCTACAGATGTGTTTCAGAAATTAATGACCATGACACAGT 253
RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
XX AAT74141;
XX
XX 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
XX *Salmonella enteritidis* 27655-3b *TnphoA* mutant agfa gene fragment.
XX
XX Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody; ds.
XX
XX

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments) 5309.115 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaataacaatctgtatc.....ctcgccaccacacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	41	9.0	7218	1 US-08-232-463-14	Sequence 14, Appl
2	34	7.5	361	1 US-08-233-768A-56	Sequence 56, Appl
3	34	7.5	456	1 US-08-233-768A-58	Sequence 58, Appl
4	32.2	7.1	1443	4 US-09-328-352-2823	Sequence 2823, Ap
5	32.2	7.1	1830121	4 US-09-557-884-1	Sequence 1, Appl
6	32.2	7.1	1830121	4 US-09-643-990A-1	Sequence 1, Appl
7	32	7.0	651	4 US-09-134-000C-1323	Sequence 1323, Ap
8	31.8	7.0	1830121	4 US-09-557-884-1	Sequence 1, Appl
9	31.8	7.0	1830121	4 US-09-643-990A-1	Sequence 1, Appl
10	31.4	6.9	1230025	4 US-09-198-452A-1	Sequence 1, Appl
11	31	6.8	1230025	4 US-09-198-452A-1	Sequence 1, Appl
12	30.6	6.7	832	4 US-09-621-976-2813	Sequence 2813, Ap
13	30.6	6.7	1350	4 US-09-328-352-3386	Sequence 3386, Ap
14	30.2	6.6	1053	4 US-09-543-681A-408	Sequence 408, App
15	30.2	6.6	1072	4 US-08-956-171E-435	Sequence 435, App
16	30	6.6	1454	2 US-08-657-392-1	Sequence 1, Appl
17	30	6.6	1454	2 US-08-657-392-1	Sequence 1, Appl
18	30	6.6	1455	2 US-08-657-392-26	Sequence 26, Appl
19	30	6.6	1455	2 US-08-657-392-26	Sequence 26, Appl
20	30	6.6	9072	4 US-08-956-171E-45	Sequence 45, Appl
21	30	6.6	92407	4 US-09-596-002-36	Sequence 36, Appl
22	29.8	6.5	640681	4 US-09-790-988-1	Sequence 1, Appl
23	29.6	6.5	474	4 US-09-621-976-18033	Sequence 18033, A
24	29.6	6.5	553	4 US-09-621-976-15491	Sequence 15491, A
25	29.6	6.5	1588	1 US-08-446-908-3	Sequence 3, Appl
26	29.6	6.5	1588	1 US-08-231-205A-3	Sequence 3, Appl
27	29.6	6.5	1588	2 US-08-871-161-3	Sequence 3, Appl

28	29.6	6.5	1589	4 US-09-023-655-1132	Sequence 1132, Ap
C 29	29.4	6.4	141	4 US-08-956-171E-2326	Sequence 2326, Ap
C 30	29.4	6.4	501	4 US-08-956-171E-654	Sequence 654, App
31	29.4	6.4	1440	4 US-09-328-352-3130	Sequence 3130, Ap
32	29.4	6.4	1710	4 US-09-543-681A-934	Sequence 934, App
33	29.4	6.4	1723	4 US-09-268-347-35	Sequence 35, Appl
34	29.4	6.4	9793	1 US-08-470-202-56	Sequence 56, Appl
35	29.4	6.4	9793	1 US-08-471-702-56	Sequence 56, Appl
36	29.4	6.4	9793	2 US-08-468-059-56	Sequence 56, Appl
37	29.4	6.4	9793	3 US-09-109-916-56	Sequence 56, Appl
38	29.4	6.4	9793	4 US-09-886-156-56	Sequence 56, Appl
39	29.4	6.4	9793	4 US-09-886-149-56	Sequence 56, Appl
40	29.4	6.4	9793	4 US-09-886-150-56	Sequence 56, Appl
41	29.4	6.4	9793	4 US-09-886-159-56	Sequence 56, Appl
42	29.2	6.4	1234	4 US-09-620-312D-953	Sequence 953, App
C 43	29.2	6.4	20284	4 US-09-526-193A-21	Sequence 21, Appl
44	29	6.4	1287	4 US-09-461-325-18	Sequence 18, Appl
45	29	6.4	1287	4 US-10-012-542-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/c
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHRIFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOULPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-9300
TELEFAX: (703) 683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO. 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZpnt-F18
US-08-232-463-14
Query Match 9.0%, Score 41, DB 1, Length 7218;

FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match
Best Local Similarity 7.1%; Score 32.2; DB 4; Length 1830121;
Matches 106; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 7 AACAAATGTTATTTATGATGTCACATATCTGGGTGCGCTGGAGTTGCACCGCGACA 66
Db 385834 ATCACTATTTGTTAAAGATTGGAATAATATGACTCTTCTCGACATCAATTGGGTT 385775
Qy 67 AATTATGATCTGCTGCTGTCAGATATATATTTGGCGTAAATGATTAAGCAAGCTTCA 126
Db 385774 AATGATGTAACCTGTTCTTCAACAATGTCATCTGCTATTCACAATATGACGTTTCA 385715
Qy 127 TTTATCAGCGCGCATATTTGTCAGTGGCGGCGGATTAATAGTCCAGATACGCCAG 186
Db 385714 CTCATATGTCAGACTATTTTGGCGTATGTTGGGTGAATGAAGTGGCAAGTCCACATTA 385655
Qy 187 GAAGATCAAACTATTTGTCGTTATTTTCAACAAGAGAGGAATATATTC 235
Db 385654 TTTAAAGCATATATGCGGCTGTTAAAGCCACAAGGGAATATTAAC 385606

RESULT 7
US-09-134-000C-1323/C
Sequence 1323 Application US/09134000C
Patent No. 6617156
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1323
LENGTH: 651
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-1323

Query Match
Best Local Similarity 7.0%; Score 32; DB 4; Length 651;
Matches 83; Conservative 0; Mismatches 65; Indels 2; Gaps 1;

Qy 217 CAGAAGAGAGAAATATTCGGGCGAAGTGCACAGGCGAGATTAATTAACCTTGGCTAT 276
Db 178 CAGAAGCTAGATTAAGCCAGTAATAATCTCCCTGTAGGCTTAATATGATTTGGCA 119
Qy 277 ATTGACAAACGGGCAATGCAAGATGCAATATATGCAAAAGCGCTTACGTAATAGT 336
Db 118 TTGGGCTAA--GGCAATCCCGCGCTTTTCACTGATGTAAGAAAGCCGATACATCTGT 61
Qy 337 GCAGCTATATTCAGAAAGCTTCTGAAT 366

Db 60 TCCCATATAGCAAGTATAGTATGATAT 31

RESULT 8
US-09-557-884-1
Sequence 1, Application US/09557884
Patent No. 6506581
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS V6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match
Best Local Similarity 7.0%; Score 31.8; DB 4; Length 1830121;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 7 AACAAATGTTATTTATGATGTCACATATCTGGGTGCGCGGATTTGCAACCGCGACA 66
Db 617140 AACACCTGTAATATATGATGTTCAATATAGTTGATGCAACCGCGCGCGACA 617199
Qy 67 AATTATGATCTGCTGCTGTCAGATATATATTTGGCGTAAATGATTAAGCAAGCTTCA 126
Db 617200 AATTAGATGATGCTTCTTCAAGCAACCATTTGACCGAGCAATTTCAAAAATAGCTT 617259
Qy 127 TTTAATCAGGC 137
Db 617260 TGTATATGGC 617270

RESULT 9
US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289
GENERAL INFORMATION:
APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White

Hamilton O. Smith
J. Craig Venter
TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186PIC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1
Query Match 7.0% Score 31.8; DB 4; Length 1830121.
Best Local Similarity 52.7%; Pred. No. 34;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;
QY 7 AACAAATTGTTATGATGTTGACATACCTGGGCGCTGGGATTGCAACCGCGACA 66
DB 617140 AACACCTGTAAATTATGTTGTTCACTCATATAGGTTGGATGCGACCGCCCCCGGACA 617199
QY 67 AATTATGATCGGCTCGCTTCAGATATATATTTGGCGTAAATGAATTAGCAAGCTTTCA 126
DB 617200 AATTACGATACCTGCTTTTCAAGCAACCATTTGACCGGACGAATTTCAAAAATACGTTT 617259
QY 127 TTTATCAGGC 137
DB 617260 TGTATATGGGC 617270
RESULT 10
US-09-198-452A-1/c
Sequence 1, Application US/09198452A
Patent No. 6559294
GENERAL INFORMATION:
APPLICANT: Griffiths, R.
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
thereof and uses thereof, in particular for the diagnosis, prevention
TITLE OF INVENTION: and treatment of infection
FILE REFERENCE: 9710-003-999
CURRENT APPLICATION NUMBER: US/09/198,452A
CURRENT FILING DATE: 1998-11-24
NUMBER OF SEQ ID NOS: 6849

SEQ ID NO 1
LENGTH: 1230025
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: misc feature
LOCATION: (1)..(15000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (15001)..(30000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (30001)..(45000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (45001)..(60000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (60001)..(75000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (75001)..(90000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (90001)..(105000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (105001)..(120000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (120001)..(135000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (135001)..(150000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (150001)..(165000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (165001)..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (180001)..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (195001)..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (210001)..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (240001)..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc feature
LOCATION: (330001)..(345000)

```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
```

```
NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

Query Match
Best Local Similarity 6.3%; Score 31.4; DB 4; Length 1230025;
Matches 59; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

```
Oy 68 ATTATGATCTGCGCTCGTCAAGATATTAATTTGGCGTAATGATTAAGCAAGTCTTCAT 127
Db 739726 ATCAAAATCTGATTTTTCAGGAATAAATATGAGGAGGTGACAGATCACTCTTGT 739667
Oy 128 TTAAATAGGCGGCATTATTTGTCAGTCGCGACGATATATG 172
Db 739666 TTGCAAGATAGGCGCATGCGCTTCAATTCATCAGGATATATG 739622
```

```
RESULT 11
US-09-198-452A-1
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Glifais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, prever
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
```

1	OTHER INFORMATION: n=a or c or g or t
2	NAME/KEY: misc.feature
3	LOCATION: (360001)..(375000)
4	OTHER INFORMATION: n=a or c or g or t
5	NAME/KEY: misc.feature
6	LOCATION: (375001)..(390000)
7	OTHER INFORMATION: n=a or c or g or t
8	NAME/KEY: misc.feature
9	LOCATION: (390001)..(405000)
10	OTHER INFORMATION: n=a or c or g or t
11	NAME/KEY: misc.feature
12	LOCATION: (405001)..(420000)
13	OTHER INFORMATION: n=a or c or g or t
14	NAME/KEY: misc.feature
15	LOCATION: (420001)..(435000)
16	OTHER INFORMATION: n=a or c or g or t
17	NAME/KEY: misc.feature
18	LOCATION: (435001)..(450000)
19	OTHER INFORMATION: n=a or c or g or t
20	NAME/KEY: misc.feature
21	LOCATION: (450001)..(465000)
22	OTHER INFORMATION: n=a or c or g or t
23	NAME/KEY: misc.feature
24	LOCATION: (465001)..(480000)
25	OTHER INFORMATION: n=a or c or g or t
26	NAME/KEY: misc.feature
27	LOCATION: (480001)..(495000)
28	OTHER INFORMATION: n=a or c or g or t
29	NAME/KEY: misc.feature
30	LOCATION: (495001)..(510000)
31	OTHER INFORMATION: n=a or c or g or t
32	NAME/KEY: misc.feature
33	LOCATION: (510001)..(525000)
34	OTHER INFORMATION: n=a or c or g or t
35	NAME/KEY: misc.feature
36	LOCATION: (525001)..(540000)
37	OTHER INFORMATION: n=a or c or g or t
38	NAME/KEY: misc.feature
39	LOCATION: (540001)..(555000)
40	OTHER INFORMATION: n=a or c or g or t
41	NAME/KEY: misc.feature
42	LOCATION: (555001)..(570000)
43	OTHER INFORMATION: n=a or c or g or t
44	NAME/KEY: misc.feature
45	LOCATION: (570001)..(585000)
46	OTHER INFORMATION: n=a or c or g or t
47	NAME/KEY: misc.feature
48	LOCATION: (585001)..(600000)
49	OTHER INFORMATION: n=a or c or g or t
50	NAME/KEY: misc.feature
51	LOCATION: (600001)..(615000)
52	OTHER INFORMATION: n=a or c or g or t
53	NAME/KEY: misc.feature
54	LOCATION: (615001)..(630000)
55	OTHER INFORMATION: n=a or c or g or t
56	NAME/KEY: misc.feature
57	LOCATION: (630001)..(645000)
58	OTHER INFORMATION: n=a or c or g or t
59	NAME/KEY: misc.feature
60	LOCATION: (645001)..(660000)
61	OTHER INFORMATION: n=a or c or g or t
62	NAME/KEY: misc.feature
63	LOCATION: (660001)..(675000)
64	OTHER INFORMATION: n=a or c or g or t
65	NAME/KEY: misc.feature
66	LOCATION: (675001)..(690000)
67	OTHER INFORMATION: n=a or c or g or t
68	NAME/KEY: misc.feature
69	LOCATION: (690001)..(705000)
70	OTHER INFORMATION: n=a or c or g or t
71	NAME/KEY: misc.feature
72	LOCATION: (705001)..(720000)
73	OTHER INFORMATION: n=a or c or g or t

```

NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (885001)..(900000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (900001)..(915000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature

```

```
Query Match      6.8%; Score 31; DB 4; Length 1230025;
Best Local Similarity 56.3%; Pred.No. 52;
Matches 58; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
```

QY 344 TTATCCAGAAAGTTCTGGAATAAAGGCCATTATTCACCGTAGCAGCAAAAACAG 403
 ||| | | | | | | | | | | | | | | | | | | | |
Db 1044632 TGATCGAAGAAGTAATCAGGCAGTGGAGTAATAATGTATAGAACATGTGTAATACAG 1044692

QY 404 CAGTGCTAGTGAGAAAAACAAGTCGACTATNGCTATTCGCCGTCCAC 446
 ||| | | | | | | | | | | | | | | | | | | | |
Db 1044692 CTGTCTAGTGAATCTCATATTGCTGTGGGCTCTTTTCGTACC 1044734

RESULT 12
US-09-621-976-2813
Sequence 2813, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jobert, S.
APPLICANT: Giordano, J.Y.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO.2813
LENGTH: 832
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:

```

; NAME/KEY: CDS
; LOCATION: 235..399
US-09-621-976-2813

```

Query Match 6.7%; Score 30.6; DB 4; Length 832;
Best Local Similarity 11.1%; Pred. No. 1.9;
Matches 21; Conservative 92; Mismatches 76; Indels 0; Gaps 0;

[illegible]

RESULT 13
US-09-328-352-3386/c
; Sequence 3386, Application US/09328352
; Patent No. 6562958

```

/ GENERAL INFORMATION:
/ APPLICANT: Gary L. Breton et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
/ TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC99-03PA
/ CURRENT APPLICATION NUMBER: US/09/328,352
/ CURRENT FILING DATE: 1999-06-04
/ NUMBER OF SEQ ID NOS: 8252
/ SEQ ID NO 3386
/ LENGTH: 1350
/ TYPE: DNA
/ ORGANISM: Acinetobacter baumannii
US-09-328-352-3386

```

Query Match	6.7%	Score 30.6;	DB 4;	Length 1350;
Best Local Similarity	65.2%	Pred. No. 2.5;		
Matches	45;	Conservative	0;	Mismatches 24; Indels 0; Gaps 0
QY	309	TATATGCCAAGCCCTTACGGTATATAGTGCAGCTATTATCCAGAAAGTTCTGGAATTA	368	
Db	586	TATATGAAAAAGCGTTGTGAGCTATAGCAGTATTTCCATTAAAGCTTCGAAAAATTA	527	
QY	369	GGCCCAATAT	377	
Db	526	TGCCACTTT	518	

```

1 RESULT 14
2 US-09-543-681A--408/C
3 ; Sequence 408, Application US/09543681A
4 ; Patent No. 6605709
5 ; GENERAL INFORMATION:
6 ; APPLICANT: GARY BRETTON
7 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
8 ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
9 ; FILE REFERENCE: 2709.1002-001
10 ; CURRENT APPLICATION NUMBER: US/09/543,681A
11 ; CURRENT FILING DATE: 2000-04-05
12 ; PRIOR APPLICATION NUMBER: US 60/128,706
13 ; PRIOR FILING DATE: 1999-04-09
14 ; NUMBER OF SEQ ID NOS: 8344
15 SEQ ID NO 408
16 ;
17 ; LENGTH: 1053
18 ; TYPE: DNA

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456

Sequence: 1 atgaaacaaatcgtatc.....ctcgccaccccaacgctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	7.7	738	US-10-369-435-3	Sequence 3, Appl1
2	34.4	7.5	18540	US-10-231-416-3	Sequence 3, Appl1
3	34.4	7.5	18540	US-10-231-426-3	Sequence 3, Appl1
4	34	7.5	529	US-10-027-632-241642	Sequence 241642,
5	33.8	7.4	203654	US-09-820-905-3	Sequence 3, Appl1
6	33.4	7.3	598	US-10-027-632-206661	Sequence 206661,
7	33.4	7.3	598	US-10-027-632-206662	Sequence 206662,
8	33.4	7.3	598	US-10-027-632-206663	Sequence 206663,
9	33.4	7.3	660	US-10-369-493-29323	Sequence 29323, A
10	33.2	7.3	535	US-10-449-857A-18	Sequence 18, Appl
11	33.2	7.3	1444	US-10-425-114-1093	Sequence 1093, Ap
12	32.8	7.2	2649	US-10-094-749-1500	Sequence 1500, Ap
13	32.4	7.1	1914	US-09-815-242-4167	Sequence 4167, Ap
14	32.4	7.1	1929	US-09-815-242-8045	Sequence 8045, Ap
15	32.2	7.1	1163020	US-10-398-221-10	Sequence 10, Appl

C	16	32.2	7.1	1830121	14	US-10-329-960-1	Sequence 1, Appl1
C	17	32.2	7.1	1830121	15	US-10-329-670-1	Sequence 1, Appl1
C	18	32.2	7.1	3011208	15	US-10-398-221-2058	Sequence 2058, Ap
C	19	32	7.0	849	15	US-10-027-632-130879	Sequence 130879, A
C	20	32	7.0	849	15	US-10-027-632-130880	Sequence 130880, A
C	21	32	7.0	849	15	US-10-027-632-130881	Sequence 130881, A
C	22	32	7.0	1353	12	US-10-282-122A-21063	Sequence 21063, A
C	23	32	7.0	1356	9	US-09-815-242-6450	Sequence 6450, Ap
C	24	32	7.0	1539	12	US-10-282-122A-25267	Sequence 25267, A
C	25	32	7.0	7640	12	US-10-424-599-51912	Sequence 51912, A
C	26	32	7.0	2614	9	US-09-070-927A-375	Sequence 375, App
C	27	31.8	7.0	933	15	US-10-027-632-14367	Sequence 14367, A
C	28	31.8	7.0	933	9	US-09-815-242-6967	Sequence 6967, Ap
C	29	31.8	7.0	933	12	US-10-282-122A-22047	Sequence 22047, A
C	30	31.8	7.0	1365	9	US-09-815-242-4243	Sequence 4243, Ap
C	31	31.8	7.0	1365	9	US-09-815-242-8196	Sequence 8196, Ap
C	32	31.8	7.0	2209	9	US-09-925-301-494	Sequence 494, App
C	33	31.8	7.0	5599	12	US-10-363-616-89	Sequence 89, Appl
C	34	31.8	7.0	1830121	14	US-10-329-960-1	Sequence 1, Appl1
C	35	31.8	7.0	1830121	15	US-10-329-670-1	Sequence 1, Appl1
C	36	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, Ap
C	37	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, Ap
C	38	31.6	6.9	1251	15	US-10-369-493-44921	Sequence 44921, A
C	39	31.6	6.9	465237	9	US-09-933-267A-1	Sequence 1, Appl1
C	40	31.4	6.9	472	9	US-09-864-761-2780	Sequence 2780, Ap
C	41	31.4	6.9	664	15	US-10-027-632-135259	Sequence 135259, A
C	42	31.4	6.9	664	15	US-10-027-632-135260	Sequence 135260, A
C	43	31.4	6.9	664	15	US-10-027-632-152059	Sequence 152059, A
C	44	31.4	6.9	808	15	US-10-027-632-160710	Sequence 160710, A
C	45	31.4	6.9	808	15	US-10-027-632-160711	Sequence 160711, A

ALIGNMENTS

RESULT 1
US-10-369-435-3
Sequence 3, Application US/10369435
Publication No. US20040002440A1
GENERAL INFORMATION:
APPLICANT: Mathews, Sarah
TITLE OF INVENTION: No. US20040002440A1 Diagnostic Agents and Uses Therefor
FILE REFERENCE: 10338-15US (2615070/VPA)
CURRENT FILING DATE: 2003-02-19
PRIOR APPLICATION NUMBER: US/10/369,435
PRIOR FILING DATE: 2000-08-18
PRIOR APPLICATION NUMBER: PCT/AU01/01021
PRIOR FILING DATE: 2001-08-17
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn version 3.2
SEQ ID NO 3
LENGTH: 738
TYPE: DNA
ORGANISM: Chlamydia pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(738)
US-10-369-435-3
Query Match 7.7%; Score 35; DB 15; Length 738;
Best Local Similarity 57.9%; Pred. No. 1.9;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
C 340 GCTATTATCCAGAAAGCTTCGAAATAGCCATATTACCCAGTACGATCCAGAA 399
D 1 GCTTTATATCAGAAATTCAGCAGTGAAGTAAATGATATGATCGTAGATATATA 60
O 400 ACACAGTTGATGTCGAGAAACAGTCGATATGCTATTCGCGTAC 446
D 61 ACACCTGTCGATGAAATGATATATGCTTGTGCTCTTTTCGTAC 107

RESULT 10
US-10-449-857A-18

	Indels	Gaps
QY	53	0
Db	207	266
QY	309	366

;; PRIOR FILING DATE: 2000-10-23
;; PRIOR APPLICATION NUMBER: 60/253,625
;; PRIOR FILING DATE: 2000-11-27
;; PRIOR APPLICATION NUMBER: 60/257,931
;; PRIOR FILING DATE: 2000-12-22
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; NUMBER OF SEQ ID NOS: 14110
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 8045
;; LENGTH: 1929
;; TYPE: DNA
;; ORGANISM: Staphylococcus aureus
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (1)...(1929)
US-09-815-242-8045

Query Match
Best Local Similarity 7.1%; Score 32.4; DB 9; Length 1929;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 263 ATAACTTGGCTATTTAGCAAGGCGCAATGCCAGTATTCGCAAGCG 322
DB 1001 AAAATTGAAATGCTTATCACTCAATTCACCAACCTATCAGTATAGAGTCTCTA 1060
QY 323 CTACGGTAAATAGTGCAGCTATTATCCGAAAGGTTCTGGAATAGGCCATATTACC 382
DB 1061 AAGCGATCATATATGCAATCATTCGGCCAAATGTATGAAATGACCTTAATTAAAA 1120
QY 383 AGTACGGTACGCAAGAAACGCGAGTT 408
DB 1121 CTATTGCTAATCAACAAAGCACTT 1146

RESULT 15

US-10-398-221-10
;; Sequence 10, Application US/10398221
;; Publication No. US20040018514A1
;; GENERAL INFORMATION:
;; APPLICANT: KUNST, Frederik
;; APPLICANT: GLASER, Philippe
;; TITLE OF INVENTION: *Listeria innocua*, genome and applications
;; FILE REFERENCE: 344 702 - US
;; CURRENT APPLICATION NUMBER: US/10/398,221
;; CURRENT FILING DATE: 2003-03-27
;; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
;; PRIOR FILING DATE: 2001-10-04
;; PRIOR APPLICATION NUMBER: FR 00/12 697
;; PRIOR FILING DATE: 2000-10-04
;; NUMBER OF SEQ ID NOS: 4025
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 10
;; LENGTH: 1163020
;; TYPE: DNA
;; ORGANISM: *Listeria innocua*
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(end)
;; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match
Best Local Similarity 7.1%; Score 32.2; DB 15; Length 1163020;
Matches 108; Conservative 0; Mismatches 108; Indels 1; Gaps 1;

QY 215 CACAAGAGAGAAATATATCGGCGCAAGTCGACAGGCGAGGAATTAACCTTGGCT 274
DB 1136649 CACAAGACATAGAGAAATGCTGTAATGCTTAATAGATTAATAATCTATTAACTG 1136708
QY 275 ATATTGACAAAGCGGCATGCAACAGATCCAGTATATCGCAAGCGCTTACGTAATA 334
DB 1136709 ATATTGACATACGGGCTATTGTAAATGAAGATAGATGTCAGGAATTAATGAGATA 1136768

QY 335 GTGACGCTATTATCCGAAAGGTTCTGGAATTAAGGCCAATATTACCAGTACGTAACG 394
DB 1136769 GT-TAGCGAATATATAGTATGCTTAAGAAAGAGTATGAAATTTAAATATACCAATGAA 1136827
QY 395 AGAAACAGCAGTTGTAGTGCAGAAACAGTCGCATAT 431
DB 1136828 CGCTACAAACCAATTTCAAGTAACTCAGGAATTT 1136864

Search completed: March 17, 2004, 08:15:08
Job time : 410.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-2
Perfect score: 456
Sequence: 1 atgaaacaacatctgtatc.....ctcgccacccaacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum March 0%
Maximum March 100%

Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hnc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_hnc:*
12: gb_estc3:*
13: gb_estc4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	132.2	29.0	220	12	B1941542 dg26906.y
2	53.6	11.8	207	9	A1368352 SMOVIFAR
3	39.4	8.6	1101	29	AL068795 Drosophila
4	39	8.6	680	12	B1618688 B1618688

5	38.8	8.5	775	9	A1664697
6	38	8.3	288	28	A2578608
7	37.8	8.3	675	12	BP007561
8	37.8	8.3	715	9	AV896333
9	37.6	8.2	601	13	B0213938
10	37.4	8.2	1214	28	CC310289
11	36.8	8.1	708	13	B0387712
12	36.8	8.1	924	28	A2529672
13	36.4	8.0	1550	28	CC211905
14	36	7.9	499	28	BH528896
15	36	7.9	579	12	BG226083
16	36	7.9	795	28	B2229556
17	36	7.9	797	28	BH505143
18	36	7.9	975	29	CNS03710
19	35.8	7.9	883	28	BH159159
20	35.6	7.8	412	10	AW140314
21	35.6	7.8	760	28	B2756658
22	35.2	7.7	659	12	BJ481788
23	35.2	7.7	791	14	CB310770
24	34.8	7.6	719	12	BG548059
25	34.8	7.6	1201	13	BX406892
26	34.6	7.6	606	10	BE581316
27	34.6	7.6	698	28	BH943841
28	34.6	7.6	700	13	BW245005
29	34.6	7.6	710	9	AV940831
30	34.4	7.5	435	14	CA856583
31	34.4	7.5	471	28	AQ055856
32	34.4	7.5	576	12	BM273661
33	34.4	7.5	577	13	BQ451444
34	34.4	7.5	581	14	CB389390
35	34.4	7.5	633	12	BJ104432
36	34.4	7.5	735	28	B2110007
37	34.4	7.5	1094	29	CNS008R5
38	34.2	7.5	666	13	B0562719
39	34.2	7.5	689	13	BQ625251
40	34.2	7.5	784	28	BZ280200
41	34.2	7.5	854	12	BM012028
42	34.2	7.5	965	9	AL549572
43	34.2	7.5	1084	28	CC239970
44	34.2	7.5	1224	14	CF238215
45	34	7.5	402	9	AW043435

ALIGNMENTS

RESULT 1
LOCUS B1941542
DEFINITION dg26906.y1 Xenopus laevis gastrula non normalized Xenopus laevis
CDNA clone IMAGE:3749963 5' similar to SW:CSGB_ECOLI P35828 MINOR
CURLIN SUBUNIT PRECURSOR. ; mRNA sequence.
B1941542
B1941542.1 GI:16256014

ACCESSION B1941542
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 220)
Clifton, S., Johnson, S. L., Blumberg, B., Song, J., Hillier, L.,
Pape, D., Martin, J., Wylie, T., Underwood, K., Theising, B., Bowers, Y.,
Peterson, B., Gibbons, M., Harvey, N., Rittler, E., Jackson, Y., McCann, R.,
Waterston, R. and Wilson, R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@atcon.wustl.edu
 Library constructed by Bruce Blumberg
 DNA Sequencing by: Washington University Genome Sequencing Center
 Source lab clone id - 3749963
 Trace considered overall poor quality This clone is available
 (image@image.llnl.gov) for further information.
 Seq primer: -40RP from Gibco
 High quality sequence stop: 1.
 Location/Qualifiers
 1..220

FEATURES
 source
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:3749963"
 /tissue_type="gastrula (stages 10.5, 11.5 mixed)"
 /lab_host="Top-10 P"
 /clone_lib="Xenopus laevis gastrula non normalized"
 /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI; cDNA was prepared from 2 ug of poly A+ RNA (equal parts from stage 10.5 and stage 11.5 gastrulae). EcoRI-XhoI cut cDNA was then ligated into UniZap-XR (Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. The library was mass excised and used to infect Top10F. Clones were picked into freezing medium (per liter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM K2HPO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4 7 H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 24 hours. Original library construction by Bruce Blumberg (Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

Query Match 29.0%; Score 132.2; DB 12; Length 220;
 Best Local Similarity 75.6%; Pred. No. 5.2e-30;
 Matches 164; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 QY CTGGTCAGAAATATATTTTGGGTAATGATTAAGCAAGTCTTGAATGACGGCG 139
 DB 3 CTAAATCATATATTAATCTTGGCGACATGATGATGATGATGATGATGATGATG 62
 QY 140 CCATTATTTGTCAGAGTCGCGACGATTAATAGTCAGAGTACGCGAGGAGATCAAAAC 139
 DB 63 CCATTAATTTGTCAGAGTCGCGACGATTAATAGTCAGAGTACGCGAGGAGCTCAAAAC 122
 QY 200 TATTTGCGCTTATTTTCAACAAGAGAGAAATATCGGCGAAGTGCAGCAGGAGGGA 259
 DB 123 TTTTGGCGTTGCTCGCAAGAGGTATGACCAACCGGCGAAGATTGACAGACGAG 182
 QY 260 ATTATTAATCTTGGTATATTTAGCAAAACGGGCAATGC 296
 DB 183 ATTATTAATCTTGGTATATTTAGTATGATGAGCGGCGAGCTCC 219

RESULT 2
 A1368352 207 bp mRNA linear EST 11-JAN-1999
 LOCUS A1368352/2
 DEFINITION SMOVMFCAR04G09SK Onchocerca volvulus microfilaria cDNA
 (SAM98MLW-OVMF) Onchocerca volvulus cDNA clone SMOVMFCAR04G09 5',
 mRNA sequence.
 ACCESSION A1368352
 VERSION A1368352.1 GI:4147105
 KEYWORDS EST.
 SOURCE Onchocerca volvulus
 ORGANISM Onchocerca volvulus
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
 Onchocercidae; Onchocerca.
 REFERENCE 1 (bases 1 to 207)
 AUTHORS Williams, S.A.
 JOURNAL Genes expressed in microfilaria of Onchocerca volvulus
 COMMENT Unpublished (1999)
 CONTACT: Steven A. Williams
 MOLECULAR Parasitology

Smith College Department of Biological Sciences
 Department of Biological Sciences, Clark Science Center, Smith
 College, Northampton, MA, 01063, USA
 Tel: 4135853826
 Fax: 4135853786
 Email: genome@smith.edu
 Seq primer: Bluescript SK.
 Location/Qualifiers
 1..207

FEATURES
 source
 /organism="Onchocerca volvulus"
 /mol_type="mRNA"
 /db_xref="taxon:6282"
 /clone="SMOVMFCAR04G09"
 /dev_stage="microfilaria"
 /lab_host="XLI-Blue MRF"
 /clone_lib="Onchocerca volvulus microfilaria cDNA (SAM98MLW-OVMF)"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Filarial nematode parasite of humans. mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kumba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 7.8 x 10⁵ independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Lizotte-Waniewski. The library is available from Dr.S.A.Williams, email:genome@smith.edu."

ORIGIN

Query Match 11.8%; Score 53.6; DB 9; Length 207;
 Best Local Similarity 91.8%; Pred. No. 1.4e-05;
 Matches 56; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 ATGAAGAAACAATGCTATTTATGATGATGATGATGATGATGATGATGATGATG 60
 DB 61 ATGAAGAAACAATGCTATTTATGATGATGATGATGATGATGATGATGATGATG 2
 QY 61 G 61
 DB 1 G 1

RESULT 3
 CNS00E93 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS CNS00E93
 DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC #
 BACR28G06 of RPCI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL068795
 VERSION AL068795.1 GI:4949039
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 1101)
 AUTHORS Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage;
 BP 191 91006 EVRY cedex - FRANCE (E-mail: seqref@genoscope.cns.fr)
 - Web: www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using a physical map of the Drosophila
 melanogaster BAC library was prepared by Kazutoyo Osoegawa and
 Aaron Mamonos in Pictet de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPCI-98 and was constructed by partial
 EcoRI digestion of Drosophila DNA provided by the BDGP from the

isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE

```
1. 1101
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR28G06"
/clone_lib="RPC1-98"
/notice="end : TET3"
```

ORIGIN

```
Query Match      8.6%; Score 39.4; DB 29; Length 1101;
Best Local Similarity 33.2%; Pred. No. 0.61;
Matches 92; Conservative 52; Mismatches 133; Indels 0; Gaps 0;

Oy 2 TGAACAAACAAATTGTTATTATGATGCTGACAAATCTGGTGGCGCTGGATGCAACG 61
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 825 KGAACAAACAACTTGTATTAATGTAATWDAAPAAASKGTCTCTTTTWTGGS 884
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Oy 62 CGACAATTAATGATCTGCTGCTGCAATATATTTGCGTAATGATTAAGCAAGT 121
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 885 TGCMTAATATATYKGGGGAAGKTAARBDISGDAKTRTATTSATSMETWATARB 944
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Oy 122 CTTCATTTAATGAGCGGCACTTATTTGTCAGTCGCGACGATTAATAGCCAGATAC 181
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 945 KSTAAATTKGSCCKKWCCTTTTBTBTBTBTBTBTBTBTBTBTBTBTBTBTBT 1004
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Oy 182 GCCGGAAGATCAAACTATGTCCTTATTTTCAAGAGAGGAATTAATGCGGGA 241
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1005 TKKARAGAGGGGGRRAAATTTGAGTAAATDAPARARAGRRARAPARARAGRGA 1064
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Oy 242 AAGTCACCGACGAGGATTAATTAATTTGCGTATAT 278
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 1065 AAWRGAAAGATADAAWTGTTTGTGRTGAWWAT 1101
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
```

RESULT 4
Bj618688/c 680 bp mRNA linear EST 01-OCT-2003
LOCUS Bj618688 NIBB Mochi; normalized Xenopus early gastrula library
DEFINITION Xenopus laevis cDNA clone XL186B22 5', mRNA sequence.
ACCESSION Bj618688
VERSION Bj618688.1 GI:37256713
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis

REFERENCE
AUTHORS Kohara,Y., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and
Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)
Kitayama,A., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and

REFERENCE
AUTHORS Kohara,Y. Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and
Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)
Kitayama,A., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and

REFERENCE
AUTHORS Kohara,Y. Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and
Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)
Kitayama,A., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and

REFERENCE
AUTHORS Kohara,Y. Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and
Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)
Kitayama,A., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and

REFERENCE
AUTHORS Kohara,Y. Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and
Xenopus laevis; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)
Kitayama,A., Teresaka,C., Mochi,M., Ueno,N., Shin-I,T. and

ORIGIN

```
Query Match      8.6%; Score 39; DB 12; Length 680;
Best Local Similarity 89.4%; Pred. No. 0.69;
Matches 42; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 410 TAGTCAGAAACAGTCGATATGCTATTCGCTGACCCAGCTAA 456
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 680 TAGTCAGAGACAGTCGCAATGCTATTCGCTGACCAACGTTAA 634
    ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
```

RESULT 5
A1664697 775 bp mRNA linear EST 07-JUL-1999
LOCUS TENG0665 T. Cruzii epimastigote normalized cDNA library Trypanosoma
DEFINITION cruzi cDNA clone n1059.r 5', mRNA sequence.
ACCESSION A1664697
VERSION A1664697.1 GI:4775685
KEYWORDS EST.
SOURCE Trypanosoma cruzii
ORGANISM Trypanosoma cruzii

REFERENCE
AUTHORS Gonzalez Rey,E., Remisz,E., Delgado Garcia,A. and Gonzalez,A.
TITLE Characterization of ESTs from Trypanosoma cruzii epimastigotes
JOURNAL Unpublished (1998)
CONTACT: Delgado Alberto
Departamento de Biología Molecular, Lab 303
Instituto de Parasitología y Biomedicina
Consejo Superior de Investigaciones Científicas C/ Ventanilla No
11, E-18001, Granada, Spain
Tel: 34 958 805058
Fax: 34 958 203523
Seq primer: T7
High quality sequence stop: 775.

FEATURES
SOURCE
1. 775
/organism="Trypanosoma cruzii"
/mol_type="mRNA"
/strain="C1 - Brenner"
/db_xref="taxon:5693"
/clone="n1059.r"
/cell_type="epimastigote"
/clone_lib="T. Cruzii epimastigote normalized cDNA library"
/note="Site_1: ECORI; Site_2: NotI; cDNA library constructed with oligo dt primed epimastigote mRNA and cloned in pT7318D phagemid with modified polylinker"

ORIGIN
Query Match 8.5%; Score 38.8; DB 9; Length 775;
Best Local Similarity 57.4%; Pred. No. 0.84;
Matches 70; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

Oy 245 TCGACGAGGAGGGAATTAATTAATTCGATATTTGACAAACGGCAATGCCAAGATG 304
 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 97 TCGACGAGGAGGGAATTAATTAATTCGATATTTGACAAACGGCAATGCCAAGATG 156
 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Oy 305 CCAGTATATGCAACAGCGCTTACGTAATAGTGACGATTAATTCAGAAAGCTTCTGAA 364
 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :
Db 157 CCTGTATATCCACACATGCTGCTGCTGACCTCAGTATTAACACAGAAAGCTGTGGA 216
 ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| : ||||| :

Oy 365 AT 366
Db 217 AT 218

RESULT 6

AZ578608/c 288 bp DNA linear GSS 08-DEC-2000
 LOCUS 25d10 Shot-gun genomic library of Rhizobium strain ANU265
 DEFINITION
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Rhizobium sp. NGR234
 Rhizobium sp. NGR234
 Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 1 (bases 1 to 288)
 Vipey, V., Rosenthal, A., Broughton, W. J. and Perret, X.
 Genetic snapshots of the Rhizobium species NGR234 genome
 Genome Biol. 1 (6), RESEARCH0014 (2000)
 21114532
 11178268
 11178268
 Contact: Virginie Vipey
 Laboratoire de Biologie Molculaire des Plantes Supérieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.vipey@bsrc.ac.uk
 Class: shotgun.
 FEATURES
 source
 1..288
 /organism="Rhizobium sp. NGR234"
 /mol_type="genomic DNA"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="25d10"
 /note="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="Vector: M13; derivative strain of NGR234 cured of
 pNGR234a"

Query Match 8.3%; Score 38; DB 28; Length 288;
 Best Local Similarity 50.0%; Pred. No. 1.1;
 Matches 95; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 QY 204 GTTCGTTATTTTCAGAGAGAGAAATATCGGGCGAAATCGACGAGGAGAAATTA 283
 DB 238 GTTCGTTATTTTCAGAGAGAGAAATATCGGGCGAAATCGACGAGGAGAAATTA 179
 QY 264 TAACCTTGCGTATTTGAGCAAGCGGCAATGCGCAAGATGATATCGCAAAAGCGC 323
 DB 178 CCAAGAGCGATACCCCTGCAAAAGGCGAAGTACAGCTTCGGGCACTTCAGTTCCG 119
 QY 324 TTACGGTAAATGTCAGCTATTTATTCAGAAAGTTTGAATAAGGCCAATTTTACCA 383
 DB 118 TTACGGCAACAATCGTGAACCCACAGAGAGAGAAATCTCTCCATCAGTCACTGCA 59
 QY 384 GTACGGTACG 393
 DB 58 GGACGTATCG 49

RESULT 7
 BP007561 675 bp mRNA linear EST 15-MAR-2002
 LOCUS BP007561 Nori Satoh unpublished cDNA library, young adult Ciona
 DEFINITION
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Ciona intestinalis
 Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 1 (bases 1 to 675)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.

TITLE
 JOURNAL
 COMMENT
 Expressed genes in Ciona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 FEATURES
 source
 1..675
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="ciad45a18"
 /tissue_type="whole animal"
 /dev_stage="young adult"
 /clone_1lb="Nori Satoh unpublished cDNA library, young
 adult"

ORIGIN
 Query Match 8.3%; Score 37.8; DB 12; Length 675;
 Best Local Similarity 50.3%; Pred. No. 1.6;
 Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
 QY 229 AATATCGGGCGAAGTGCACGAGGGAATTAATCTTGGCTATTTGAGCAACG 288
 DB 257 AATCTTTAATGAAAGAAATGTCAGAACTAAACCTGTACTGAAACAGGAAAAA 316
 QY 289 GGCATGCGCAAGATGCGCATATATCCCAAGCGCTTACGGTAATAGCTATTAATC 348
 DB 317 TGCTCCACCTCGCAGGCAATGTATTAATCAAAAAGCAATTTAGCTTGAACCT 376
 QY 349 CAGAAAGTCTGGAATTAAGCCCAATTAATCCAGTACGTAAGCAAGAAACAGCAGTT 408
 DB 377 CAATTAATTTCTGCAATTAATGACAAATTCACAGATGATTAACACACTACACAGAT 436
 QY 409 GTAGT 413
 DB 437 GAACCT 441

RESULT 8
 AV896333/c 715 bp mRNA linear EST 09-NOV-2001
 LOCUS AV896333 Nori Satoh unpublished cDNA library, young adult Ciona
 DEFINITION
 ACCSSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 EST.
 Ciona intestinalis
 Ciona intestinalis
 Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
 Phlebobranchia; Clonidae; Ciona.
 1 (bases 1 to 715)
 Satoh, N., Satou, Y., Kohara, Y. and Shin-I, T.
 Expressed genes in Ciona intestinalis
 Unpublished (2000)
 Contact: Nori Satoh
 Department of Zoology
 Kyoto University
 Sakyo-Ku, Kyoto, Kyoto 606-8502, Japan
 Tel: 81-75-753-4081
 Fax: 81-75-705-1113
 Email: satoh@ascidian.zool.kyoto-u.ac.jp.
 FEATURES
 source
 1..715
 /organism="Ciona intestinalis"
 /mol_type="mRNA"
 /db_xref="taxon:7719"
 /clone="ciad45a18"
 /tissue_type="whole animal"
 /dev_stage="young adult"

/clone_lib="Nori Satoh unpublished cDNA library, young adult"

Query Match 8.3%; Score 37.8; DB 9; Length 715;
Best Local Similarity 50.3%; Pred. No. 1.7;
Matches 93; Conservative 0; Mismatches 92; Indels 0; Gaps 0;

229 AATATGGGGGAAGTGCACGAGGAGGAAATTATCTTGCTATATTGAGCAACG 288
DB |||||
620 AATCTTCAATGAAAAAATGTCAGAACTAAACCTGTACTGAAACAGAAAAA 561
QY 289 GCAATGCCAAGATGCGATATATGCAAAAGCGCTTACGTAATAGTCAAGCTATTATC 348
DB TGCTCCACTCTGCGAGGCAATGTATATAGTACAAAAGAAAGCAATTAGCTTTGAACT 501
QY 349 CAGAAAGTTCTGGAATTAAGGCGCAATATTACCCAGTACGTAACGACAGAAAAACAGACTT 408
DB 500 CAATTAATTTCTGCAATATATGACAAATTTCAGAGATGATTAACGACACTACACAGAT 441
QY 409 GTAGT 413
DB 440 GAATC 436

RESULT 9
LOCUS BU213938 601 bp mRNA linear EST 25-NOV-2002
DEFINITION 603108484F1 CSEQCHN04 Gallus gallus cDNA clone CHESTS1g22 5', mRNA sequence.

ACCESSION BU213938
VERSION BU213938.1 GI:25389321
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus

REFERENCE 1 (bases 1 to 601)
AUTHORS Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R., Wilson, S.A. and Hubbard, S.J.
TITLE A Comprehensive Collection of Chicken CDNA
JOURNAL Curr. Biol. 12 (22), 1965-1969 (2002)
MEDLINE 22335534
PUBMED 12445392

COMMENT

Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
(UMIST)
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

Location/Qualifiers

FEATURES

source

1..601
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CHESTS1g22"
/cissue_type="whole embryo"
/dev_stage="20-21"
/lab_host="DH10B"
/clone_lib="CSEQCHN04"
/note="Organ: whole embryo; Vector: pBluescript II KS(+);
Site 1: EcoRI; Site 2: NotI; This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dt) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunt-ended, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the

ORIGIN

Query Match 8.2%; Score 37.6; DB 13; Length 601;
Best Local Similarity 50.6%; Pred. No. 1.8;
Matches 91; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 87 AGAATTAATTTTCCGGTAATTAATTAAGCAAGCTTCATTATACAGCGGCATTAT 146
DB 172 AGCAAAAATTTAGCGTAATTAATGACAGACACACCTGAGCTACCTAGGCGCTTAT 231
QY 147 TGCTCAAGTCGCGACGGATTAATAGTCCAGATACGCCAGAGATCAAACTATTGTC 206
DB 232 CAGTCCCAAGCCAAAGGAAGGGTTTGCCACCTGATTCAGAAAAGATGAAAGAGTGC 291
QY 207 CGTTATTTCACAAAGAGGAATTAATCGGCGCAAAAGTGCACCAAGCGGAATTATTA 266
DB 292 CAGTCTTCTTCTGTGATGACGTAAATTAAGTAAGGCGCTATGAAACGCAATTCTGA 351

RESULT 10
LOCUS CC310289/c 1214 bp DNA linear GSS 14-MAY-2003
DEFINITION TAM32-32B18, ECL 1 TAM32 Gallus gallus genomic clone TAM32-32B18,
genomic survey sequence.

ACCESSION CC310289
VERSION CC310289.1 GI:30703437
KEYWORDS GSS
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus

REFERENCE 1 (bases 1 to 1214)
AUTHORS Kremitzki, C., Higginbotham, J., Wylie, K., Carter, J., McPherson, J.,
Warren, W., Graves, T., Mardis, E. and Wilson, R.
TITLE Gallus gallus BAC End Reads.
JOURNAL Unpublished (2003)
COMMENT Contact: Richard K. Wilson

Genome Sequencing Center
Washington University School of Medicine
Email: submissions@wustl.edu
Insert Length: 152000 Std Error: 0.00
Seq primer: ECL1 TACGACTCACTATTAAGGCGC
Class: BAC ends

High quality sequence start: 20
High quality sequence stop: 781.

Location/Qualifiers

FEATURES

source

1..1214
/organism="Gallus gallus"
/mol_type="genomic DNA"
/strain="Red Jungle Fowl"
/db_xref="taxon:9031"
/clone="TAM32-32B18"
/sex="female"
/cell_line="UCDD001, indred 256"
/clone_lib="TAM32"
/note="Vector: pECBAC1; Site 1: EcoRI; Site 2: EcoRI;
TAM32 Female Chicken Library - for library and clone
ordering information: <http://www.hbz.tamu.edu>"

ORIGIN

Query Match 8.2%; Score 37.4; DB 28; Length 1214;
Best Local Similarity 47.3%; Pred. No. 2.6; Indels 126; Gaps 0;
Matches 113; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 71 ATGATCGGCTCGTTCAGAAATATATTTTGGGTAAATGAATTAAGCAAGCTTCATTTA 130
DB 1027 AATATTTGATGTTTCAAGTAATATTTTGAAGGGGCAATTACAGATGGAAGTTTCAAAA 968

QY 131 ATCAGCGCCATATTTGCTCAATCGGACGAGTAATAGTCCAGAGTACGCCAGAG 190
 DB 967 ATCCAGGGAACTTTATTTAGTGTAGAGAACCATCTCGTTAATTCCTGGCTTTGAGT 908
 QY 191 GATCAAACTATTTGCTCGTATTTTCAAGAGAGAGAAATATATGGGCGAAAGTCCGCC 250
 DB 907 TAGCAAAATGGATACCTTAAAGAGAACAAAGAAACCTGGCCATTTCGGTCTCCCAA 848
 QY 251 AGCGAGGATATTAATCTTTGGCTATTTAGCAAAACGGGCAATGCCAAGATGCCAGT 309
 DB 847 TACCGAGAAACAATATTTATCTATCTCTCAACAGCTTCAAGAGAGCTAATCTAT 789

RESULT 11

BU387712 708 bp mRNA linear EST 28-NOV-2002
 LOCUS 603857374F1 CSEQCHN75 Gallus gallus cDNA clone CHEST863c20 5', mRNA
 DEFINITION

ACCESSION BU387712
 VERSION BU387712.1 GI:25895713
 KEYWORDS EST
 SOURCE Gallus gallus (chicken)
 ORGANISM Gallus gallus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
 Phasianinae; Gallus

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (bases 1 to 708)
 Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
 Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A., and Hubbard, S.J.
 A Comprehensive Collection of Chicken cDNAs
 Curr. Biol. 12 (22), 1965-1969 (2002)
 1245392
 Contact: Simon Hubbard
 Department of Biomolecular Sciences
 University of Manchester Institute of Science and Technology
 (UMIST)
 PO Box 88, Manchester, M60 1QD, UK
 Tel: 01612008930
 Fax: 01612360409
 Email: Simon.Hubbard@umist.ac.uk

FEATURES

source

1..708
 Location/Qualifiers
 /organism="Gallus gallus"
 /mol_type="mRNA"
 /strain="White Leghorn, Hisex"
 /db_xref="taxon:9031"
 /clone="CHEST863c20"
 /dev_stage="36"
 /lab_host="DH10B"
 /clone_id="CSEQCHN75"
 /note="Organ: trunks; Vector: pBluescript II KS(+);
 Site_1: EcoRI; Site_2: NotI; This normalized library was
 constructed from 1 million independent clones. cDNA
 synthesis was initiated using an oligo(dT) primer, using
 methylated C in the first strand synthesis reaction.
 Following this first strand reaction, double-stranded cDNA
 was blunt-ended, ligated to NotI adapters, digested with
 EcoRI, size-selected, and cloned into the NotI and EcoRI
 compatible sites of a custom modified MCS of the
 pBluescript (KS+) vector. The library was normalized in 2
 rounds using conditions adapted from Soares et al., PNAS
 (1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
 (1996): 791, except that a significantly longer
 reannealing hybridization was used."

ORIGIN

Query Match 8.1%; Score 36.8; DB 13; Length 708;
 Best Local Similarity 50.6%; Pred. No. 3.4;
 Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 89 AATATATTTTGGCGTAATGAATTAAGCAAGTCTTCAATTAATCAAGCGGCCATTATTG 148

DB 17 AAGAGATCTACCCGTAATATGACAGAGACCAAGCTGAGCTGACTGACCTATGATCA 76
 QY 149 GTCAAGTCGACAGATTAATAGTCCAGAGTACCGCAGAGAGATCAATATTTCCG 208
 DB 77 GTCCCAAGCCAAAGAAAGGCTTTGCCACTGATTTCAGAAAGAGTGAAGAGGTGCCA 136
 QY 209 TTAATTCACAAAGAGAGAAATATATCGGCGAAGTGCACAGCAGGAAATTAT 264
 DB 137 GTCTCTTCTGATGTGACGTAATATTAAGTGAAGGCTATATGAAAGCGCAATTTT 192

RESULT 12

AZ529672 924 bp DNA linear GSS 03-NOV-2000
 LOCUS ENTBM10TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION AZ529672
 VERSION AZ529672.1 GI:11082756
 KEYWORDS GSS

SOURCE

Entamoeba histolytica
 Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE

1 (bases 1 to 924)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished (2000)

AUTHORS

Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@igr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library

Seq primer: M3-Reverse
 Class: Shotgun
 High quality sequence start: 33
 High quality sequence stop: 745.
 Location/Qualifiers

FEATURES

source

1..924
 Location/Qualifiers
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_id="Entamoeba histolytica Sheared DNA"
 /note="Vector: pBOS1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Borell, Oxford University Press, 1999)."

ORIGIN

Query Match 8.1%; Score 36.8; DB 28; Length 924;
 Best Local Similarity 50.6%; Pred. No. 3.7;
 Matches 89; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 73 GATCGGCTCGTTCAGATATATTTTGGGTAATGATTAAGCAAGTCTTCAATTAAT 132
 DB 740 GATGTACACTTAAGAATTTTATTAAGGGGTAAACATTTAGGGGTATTTACATTTGAA 799
 QY 133 CAGCGGCAATTAATTTGCTCAAGTGGGACGATATATAGTGCACAGTACGCAAGAGA 192

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 499)	Whole genome shotgun sequencing of <i>Brassica oleracea</i>	unpublished (2001)	
2	Other: GSSS: BOGPO61TF	Contact: Chris Town		
3	TIGR	9712 Medical Center Drive, Rockville, MD 20850, USA.		
4	Tel: 301-838-3523			
5	Fax: 301-838-0208			
6	Email: cdtown@tigr.org			
7	DNA is from a doubled haploid provided by Tom Osborn.			
8	Seq primer: TR			
9	Class: sheared ends.			
10	Location/Qualifiers			
11	1..499			
12	/organism="Brassica oleracea"			
13	/mol_type="genomic DNA"			
14	/strain="TO1000DH3"			
15	/db_xref="taxon:3712"			
16	/clone_id="BOGPO61"			
17	/note="Vector: pHOSt; Site: 1: BstXI; 2-3 kb sheared genomic DNA inserted into pHOSt using BstXI linkers"			
18	Query Match	7.9%; Score 36; DB 28; Length 499;		
19	Best Local Similarity	46.4%; Pred. No. 5.4;		
20	Matches 117; Conservative	0; Mismatches 135; Indels 0; Gaps 0;		
21	QY	183 CCAGGAAGCATCAAACTATTTCCTTATTTTCACAGAAAGAGAAATATCGGGCGAA	242	
22	DB	470 CCGGAAACGAGCAATGATTCACGATTTGAAGATGACAGGGCAATGCTATCA	411	
23	QY	243 AGTCACGACGAGGAAATTAATTAATCTTGGCTATTTAGACAAACGGGCAATGCCAAGA	302	
24	DB	410 AGAAGATCAATAGGAGAGATGATCATTTGATGATTTTTCATGAGCTATTTCAAGACTATTTGA	351	
25	QY	303 TGCAGATATTCGAAAGCGCTTAACGCTATTAAGTCAGCTATTTTCAGAAAGGTTCTGG	362	
26	DB	350 GGGCAACGAGAGAAACCGTAATGCTGCTTATCAACAAATGCTGAGCGCAGAAAGAA	291	
27	QY	363 AATTAAGGCAATTTATCCAGTACGCTGACGAGAAACAGAGTTGATGACGAAGA	422	
28	DB	290 CGAGAGCTATTTACATGATCCCAAAAGCTGACAGATCAAGAGGCGCTATTTCTATTTCA	231	
29	QY	423 GTGCGATATGCGC	434	
30	DB	230 TGCCGACAAAGC	219	
31	RESULT 15			
32	LOCUS	BG226083	579 bp	mRNA linear EST 09-MAY-2001
33	DEFINITION	K080h12.y1 TBN95TM-SSR Strongyloides stercoralis cDNA 5' similar to		
34	ACCESSION	WP_26686.3	CE00457.1	mRNA sequence.
35	VERSION	BG226083		
36	KEYWORDS	BG226083.1	GI:12713638	
37	SOURCE	EST.		
38	ORGANISM	Strongyloides stercoralis		
39	REFERENCE	Strongyloides stercoralis		
40	AUTHORS	Eukaryote; Metazoa; Nematoda; Chromadorea; Rhabditida;		
41		Panagrolaimidae; Strongyloidae; Strongyloides.		
42		1 (bases 1 to 579)		
43		McCartier, J., Clifton, S., Chapel, B., Pape, D., Martin, J.,		
44		Wyle, T., Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B.,		
45		Bowers, Y., Gibbons, R., Ritter, B., Bennett, J., Franklin, C.,		
46		Tsagaris, R., Ronko, I., Kennedy, S., Maguire, L., Beck, C.,		
47		Underwood, K., Stepcoe, M., Allen, M., Peterson, B., Shaller, T.,		
48		Harvey, N., Schurk, R., Kohn, S., Shin, T., Jackson, Y., Cardenas, M.,		
49		McCann, R., Waterson, R. and Wilson, R.		

TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH (nutman@nih.gov). DNA Sequencing by: Washington University Genome Sequencing Center St. Louis.
High quality sequence stop: 417.
Location/Qualifiers

FEATURES
SOURCE

1..579
/organism="Strongyloides stercoralis"
/mol_type="mRNA"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/lab_host="XL-1 Blue MRF" (Stratagene)"
/clone_lib="TBN95RW-SSR"
/note="Vector: Lambda Uni-ZAP XR (Stratagene), Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."

ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 579;
Best Local Similarity 49.0%; Pred. No. 5.6;
Matches 96; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

QY 7 AACAAATTGTATTATGATGTCACAACTACTGGGTCGGCTGGGATTGCAACCGCACA 66
   |||||
Db 271 AAAAACTTATTATTGTTATGTTGATTATGACAGATGCTCCAGATTTTGCTTCATA 330
   |||||
QY 67 AATTATGATCTGGCTGCTTCAGAAATATATTTTGGGTAATGAAATAGCAAGCTTCA 126
   |||||
Db 331 AATCTTATACAGCTCCGCACTTATACATTTTCACGTAAGGATCTAGGAATCTGAT 390
   |||||
QY 127 TTTAATCAGGGCGCATTTATGTCAGAGTCGCGACGATTAATAGTCCAGATCGCCAG 186
   |||||
Db 391 GATTCAAATGATGTTCAAAATTAATGATTTGAAGCTGATGTTTGGCAAGATTTGTTAAA 450
   |||||
QY 187 GAAGATCAAAACTAT 202
   |||||
Db 451 GAAGAACTGAAATAT 466

```

Search completed: March 16, 2004, 04:28:38
Job time : 2238.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds
(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaacttttaaaagtaga.....cgaccgtcactgactactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_bhg:*

3: gb_in:*

4: gb_ov:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_ests:*

12: gb_ey:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_ests:*

28: em_un:*

29: em_vl:*

30: em_hhg_hum:*

31: em_hhg_inv:*

32: em_hhg_other:*

33: em_hhg_mus:*

34: em_hhg_pln:*

35: em_hhg_rtd:*

36: em_hhg_mam:*

37: em_hhg_vrt:*

38: em_hhg_vrt:*

39: em_hhg_hum:*

40: em_hhg_mus:*

41: em_hhg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	6	AX814811
2	456	100.0	648	1	ECOCGAA
3	454.4	99.6	4680	1	ECOCGABDG
4	454.4	99.6	10346	1	AE000205
5	454.4	99.6	15047	1	D90741
6	395	86.6	1711	1	AE275733
7	395	86.6	10190	1	AE005315
8	395	86.6	327773	1	AP002554
9	382.2	83.8	306358	1	AE016759
10	340.8	74.7	10370	1	AE015131
11	340.8	74.7	292504	1	AE016981
12	311.4	68.3	437	1	AF237726
13	259.2	56.8	5103	1	STAJ2301
14	259.2	56.8	22411	1	AE008749
15	257.6	56.5	254050	1	AL627269
16	257.6	56.5	301983	1	AE016840
17	254.4	55.8	456	6	144909
18	254.4	55.8	2067	1	SEU43280
19	253.6	55.6	1048	1	STAGFBA
20	252.8	55.4	2883	1	ESAS15702
21	246.4	54.0	2889	1	CSP515700
22	230.8	50.6	2920	1	CPR515701
23	186	40.8	19201	1	D90742
24	173.4	38.0	361	6	144908
25	121.8	26.7	230	1	SEU53207
26	83.8	18.4	1212	1	EC0131756
27	78	17.1	78	6	AX814809
28	72	15.8	72	6	AX814798
29	66	14.5	66	6	AX814808
30	43.4	9.5	301214	1	AE016786
31	43.2	9.3	85600	9	AC138647
32	42.6	9.3	135259	9	HS127820
33	42.6	9.3	165074	9	AC093519
34	41.6	9.1	181047	2	AC016765
35	41.6	9.1	198599	9	AC108448
36	41.2	9.0	159070	2	AC128114
37	41.2	9.0	186562	2	AL451079
38	41.2	9.0	200368	9	AC096541
39	41.2	9.0	241394	2	AC132789
40	40.8	8.9	165074	9	AC093519
41	40.6	8.9	236533	2	AC103433
42	40.2	8.8	183458	2	AC129554
43	40	8.8	180995	9	AC117500
44	40	8.8	187615	2	AC020724
45	39.8	8.7	999	1	AY077581

ALIGNMENTS

RESULT 1

AX814811

LOCUS

DEFINITION Sequence 15 from Patent WO03064446.

ACCESSION AX814811

VERSION AX814811.1 GI:39104001

KEYWORDS

SOURCE

ORGANISM

Escherichia coli

Bacteriophages

Escherichia coli

Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Escherichia.

REFERENCE

1 Bioerck, L., Olsen, A., Wikstrom, M. and Herwald, H.

Peptides

PATENT: WO 03064446-A 15 07-AUG-2003;

JOURNAL

FEATURES

source

1..456
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAE83541.1"
/db_xref="GI:39104002"
/translation="MKLKYKALIAIVFSGSALAVPQYGGGNGHGGGNSGPNSE
LDQNGKXSEMTVKQFGGNGAAYDQTAASNSVVTQVGFNNATAHQY"

CDS

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 456; DB 6; Length 456;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTAGAAGCAATTCGACCAATTCGATTCCTCGGTAGCGCTTGGA 60
DB 1 ATGAACTTTTAAAGTAGAAGCAATTCGACCAATTCGATTCCTCGGTAGCGCTTGGA 60
QY 61 GGTGTTCTCTCACTACGCGCGCGCGGTAAACCAACCGGTGATGCGGTATTAAGCGC 120
DB 61 GGTGTTCTCTCACTACGCGCGCGCGGTAAACCAACCGGTGATGCGGTATTAAGCGC 120
QY 121 CCAAAATTCGAGCTGAACATTTACAGTAGCGGTGCGGTAACTCTGCACTTCTGCA 180
DB 121 CCAAAATTCGAGCTGAACATTTACAGTAGCGGTGCGGTAACTCTGCACTTCTGCA 180
QY 181 ACTGATGCCGTAACTCTGACTGATTAATCCAGCATGCGCGCGGTAAATGTCAGAT 240
DB 181 ACTGATGCCGTAACTCTGACTGATTAATCCAGCATGCGCGCGGTAAATGTCAGAT 240
QY 241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATCGAACCGGCTTGCGTAAACGC 300
DB 241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATCGAACCGGCTTGCGTAAACGC 300
QY 301 GCTACTCTTGATCAGTGGAACGGCAAAATTCGAAATGAGCGTTAAACAGTTGAGT 360
DB 301 GCTACTCTTGATCAGTGGAACGGCAAAATTCGAAATGAGCGTTAAACAGTTGAGT 360
QY 361 GCGAAGCGTGTGACAGTTGACAGCTGATCTAATCTCCGTCACAGTCAAGTT 420
DB 361 GCGAAGCGTGTGACAGTTGACAGCTGATCTAATCTCCGTCACAGTCAAGTT 420
QY 421 GCGTTGTGTAACAGCGACCGCTCATCATGACTTAA 456
DB 421 GCGTTGTGTAACAGCGACCGCTCATCATGACTTAA 456

RESULT 2

LOCUS

ECOCSSGA

648 bp DNA

DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
VERSION L04979.1 GI:290424
KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
SOURCE Escherichia coli
ORGANISM Escherichia coli

REFERENCE

AUTHORS

TITLE

Olsen, A., Arngvist, A., Hammar, M., Sukupolvi, S. and Normark, S.
The RpoS sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli
in Escherichia coli

JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)
MEDLINE 93211294
PUBMED 8459772

COMMENT

On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara) DNA.
Location/Qualifiers

FEATURES

source

1..648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/issue="11"
/gene="csgA"
83..538
/gene="csgA"
83..538
/note="precursor protein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="curlin subunit"
/protein_id="AAA23616.1"
/db_xref="GI:290425"
/translation="MKLKYKALIAIVFSGSALAVPQYGGGNGHGGGNSGPNSE
LDQNGKXSEMTVKQFGGNGAAYDQTAASNSVVTQVGFNNATAHQY"

CDS

ORIGIN

Query Match
Best Local Similarity 100.0%; Score 456; DB 1; Length 648;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

sig_peptide
mat_peptide

/gene="csgA"
/evidence=experimental
143..535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

QY 1 ATGAACTTTTAAAGTAGAAGCAATTCGACCAATTCGATTCCTCGGTAGCGCTTGGA 60
DB 83 ATGAACTTTTAAAGTAGAAGCAATTCGACCAATTCGATTCCTCGGTAGCGCTTGGA 142
QY 61 GGTGTTCTCTCACTACGCGCGCGGTAAACCAACCGGTGATGCGGTATTAAGCGC 120
DB 143 GGTGTTCTCTCACTACGCGCGCGGTAAACCAACCGGTGATGCGGTATTAAGCGC 202
QY 121 CCAAAATTCGAGCTGAACATTTACAGTAGCGGTGCGGTAACTCTGCACTTCTGCA 180
DB 203 CCAAAATTCGAGCTGAACATTTACAGTAGCGGTGCGGTAACTCTGCACTTCTGCA 262
QY 181 ACTGATGCCGTAACTCTGACTGATTAATCCAGCATGCGCGCGGTAAATGTCAGAT 240
DB 263 ACTGATGCCGTAACTCTGACTGATTAATCCAGCATGCGCGCGGTAAATGTCAGAT 322
QY 241 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATCGAACCGGCTTGCGTAAACGC 300
DB 323 GTTGTCAGGAGCTCAGATGACAGCTCAATCGATCGAACCGGCTTGCGTAAACGC 382
QY 301 GCTACTCTTGATCAGTGGAACGGCAAAATTCGAAATGAGCGTTAAACAGTTGAGT 360
DB 383 GCTACTCTTGATCAGTGGAACGGCAAAATTCGAAATGAGCGTTAAACAGTTGAGT 442
QY 361 GCGAAGCGTGTGACAGTTGACAGCTGATCTAATCTCCGTCACAGTCAAGTT 420
DB 443 GCGAAGCGTGTGACAGTTGACAGCTGATCTAATCTCCGTCACAGTCAAGTT 502
QY 421 GCGTTGTGTAACAGCGACCGCTCATCATGACTTAA 456
DB 503 GCGTTGTGTAACAGCGACCGCTCATCATGACTTAA 538

RESULT 3

BCCSGABDG

LOCUS ECCSABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION E.coli csdg, csgf, csge, csgd, csgeB, csgeA, and orfC genes.
ACCESSION X90754
VERSION X90754.1 GI:1147558
KEYWORDS csge gene; csgeB gene; csge gene; csge gene; csge gene;
orfC gene.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1
AUTHORS Hammar,M., Arundvist,A., Bian,Z., Olsen,A. and Normark,S.
TITLE Expression of two csg operons is required for production of
fibronectin- and congo red-binding curli polymers in Escherichia
coli K-12
JOURNAL Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE 96414468
PubMed 8817489
REFERENCE 2 (bases 1 to 4680)
AUTHORS Hammar,M.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
SWEDEN
FEATURES
source Location/Qualifiers
1..4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:562"
/map="23.15 minutes"
1..3140
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="MC4100"
/db_xref="taxon:562"
3141..4680
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="W3110"
/db_xref="taxon:562"
<1..>4680
/insertion_seg="182 (partial)"
complement(133..966)
/gene="csge"
complement(133..966)
/gene="csge"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
protein_id="CAA62277.1"
/db_xref="GI:1147559"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORFLVAVMLSGCCTAPKRAAPRTIMPRAQSYKDLTHLPA
PTGKTFPSVYNIQDETQCFKPYPAASNSTANPOSATMLATLKDSRFITLEQGLQ
NLNLRKTIKRAQENGTVAINNRIPLQSLTANITVESIIIGESNVSGGVARYRG
IGADTQYLDQDIANLRAVNVTGBILSVNTSKTILSYEQAGVFRIIDYORLLEB
VGYSNEPVMCLMSALETGYVFLINDIDRLMDLQKAKERNDIIVKRYHMSVPE
S"
complement(993..1409)
/gene="csge"
complement(993..1409)
/gene="csge"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
protein_id="CAA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translation="MRVXHAVVLLMLISPLSMAGTMTPQFNPNFGNPNNGAFLNLS

gene
complement(1343..1823)
/gene="csge"
complement(1343..1823)
/gene="csge"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
protein_id="CAA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translation="MKRYLKIYVAERLFAAGNLAHVEVPEGLTHTTSSIGHDFY
RAFSDKWSEDYTGLTINERPSARWSMTITVQDVIPTFLPPLKDFEKTFFAL
IQTEALNRROINQALSTGDLAHDF"
complement(1828..2478)
/gene="csge"
complement(1828..2478)
/note="necessary for transcription of the csgeA operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
protein_id="CAA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MPNEVSHIGHLTLLITKSSLOQATALLQHLKQSLATIGKLANIQ
RSLDDISSGIILLDMEMADKLIHYQDTLSRNNNIKILLNTPEDYPRDIENP
HINGVFPYSMQDERVNVQGVLRGEVFPQKLASYITSGNRYNSTSALLTHRE
KEILNKRIRIGASNNENRSLPISENTYKTHLYLFKIAYAKRTQANVANDNLR"
complement(2626)
/note="transcriptional start site for csgeB operon"
3140
/note="transcriptional start site for csgeB operon"
3233..3688
/gene="csgeB"
3233..3688
/gene="csgeB"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
protein_id="CAA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKKLLFMMULTIIGAGGIAAAGYDLANSEYFANVELSKSFN
QAAIIQAGTNNSSAQLRGGSKLLAAVVAQSSNRRAKIDQDGVNLAYIDQAGSANDA
SISQAGATNTAMIIQKSGSKNANITQYGTQKTAIVVROSQMAIRVQR"
3729..4184
/gene="csgeA"
3729..4184
/gene="csgeA"
/note="major subunit"
/codon_start=1
/transl_table=1
/product="curlin/CsgA protein"
protein_id="CAA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="WKLLKVAIATAIYFSGSALAGVVPYGGGNGHGGGNNSCPSNE
LNIYQGGGNSALALQTDARNSDLTITQHGCGAGADVGQSDDSIDLTRGFGNSAT
LDQWNGNSSEMTYVQFGGGAADVQPAASNSVNVTVQFGNNATAHQY"
3729..3788
/gene="csgeA"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/transl_table=1
sig_peptide
gene
CDS

/product="putative curli production protein"
 /protein_id="CA62283.1"
 /db_xref="GI:1147565"
 /translation="MTLLLLAASSLTNTTQGGDYTTIPEVTLTQSLCRVQL
 SLRESGSGSQTKRKETLSLPANQPIALTKLSLNPDRKIVTVTSDQSILHLSQ
 WPSSEKS"

ORIGIN

Query Match 99.6%; Score 454.4; DB 1; Length 4680;
 Best Local Similarity 99.8%; Pred. No. 3,5e-121;
 Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAGAACCAATTCACCAATTCGATTTCTCGGTACCGCTTGGCA 60
 DB 3729 ATGAACCTTTAAAGTAGAACCAATTCACCAATTCGATTTCTCGGTACCGCTTGGCA 3788
 QY 61 GGTGTTGTTCTCTCACTACGCGCGCGCGCGGTAAACACCGGTGGTGGCGGTATATACCGGC 120
 DB 3789 GGTGTTGTTCTCTCACTACGCGCGCGCGCGGTAAACACCGGTGGTGGCGGTATATACCGGC 3848
 QY 121 CCMAATTCAGCTGAACATTTACCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 180
 DB 3849 CCMAATTCAGCTGAACATTTACCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 3908
 QY 181 ACTGATGCCGTTAACTTGAATTCATTTACCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 3908
 DB 3909 ACTGATGCCGTTAACTTGAATTCATTTACCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 3968
 QY 241 GTTGTCAGGCGCTCAGATGACGCTCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 300
 DB 3969 GTTGTCAGGCGCTCAGATGACGCTCAATTCAGTACGCGGTAACTTCTGCACTTGTCTGCA 4028
 QY 301 GCTACTCTTATATAGTGAACCGCAAAATTCGAAATGACGCGGTAACTTCTGCACTTGTCTGCA 360
 DB 4029 GCTACTCTTATATAGTGAACCGCAAAATTCGAAATGACGCGGTAACTTCTGCACTTGTCTGCA 4088
 QY 361 GCGAACGCGTCTCAGTACGCGGTAACTTCTGCACTTGTCTGCACTTGTCTGCACTTGTCTGCA 420
 DB 4089 GCGAACGCGTCTCAGTACGCGGTAACTTCTGCACTTGTCTGCACTTGTCTGCACTTGTCTGCA 4148
 QY 421 GCGTTTGTGAACACGCGCGGTAACTTCTGCACTTGTCTGCACTTGTCTGCACTTGTCTGCA 456
 DB 4149 GCGTTTGTGAACACGCGCGGTAACTTCTGCACTTGTCTGCACTTGTCTGCACTTGTCTGCA 4184

RESULT 4
 AEO00205 10346 bp DNA linear BCT 01-DEC-2000
 LOCUS Escherichia coli K12 MG1655 section 95 of 400 of the complete
 DEFINITION genome.
 ACCESSION AEO00205 U00096
 VERSION AEO00205.1 GI:1787265
 KEYWORDS
 SOURCE
 ORGANISM Escherichia coli K12
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 10346)
 Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
 Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
 Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
 Mu, B. and Shao, Y.
 The complete genome sequence of Escherichia coli K-12
 Science 277 (5331), 1453-1474 (1997)

TITLE JOURNAL
 MEDLINE 97428617
 PUBMED 9278503
 REFERENCE 2 (bases 1 to 10346)
 Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:

REFERENCE

608-263-7459
 3 (bases 1 to 10346)
 Blattner, F.R.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
 608-263-7459

AUTHORS

Plunkett, G. III.
 TITLE Direct Submission
 JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

COMMENT

This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES

source
 1..10346
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(1..48)
 /note="IS3"
 <1..5
 /note="factor Sigma70; predicted +1 start at 1094689"
 5..32
 /note="factor Sigma70; predicted +1 start at 1094716"
 69..392
 /gene="b1028"
 69..392
 /gene="b1028"
 /function="orf; Unknown"
 /note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
 to 58 residues of an approx. 1888 aa protein PAS2_CANAL
 SW: P43098"
 /codon_start=1
 /transl_table=1
 /product="orf; hypothetical protein"
 /protein_id="AAC74113.1"
 /db_xref="GI:1787266"
 /translation="MEDGVQAMRDYLAGLDIASPEHVLNVTAKSEVAPSLIKENTLS
 LHLTHYKWTESLDPLFMPTPVAFLISNKPYLGMNLDPAGVDDQRVWRCKKAFSD
 AKVKK"
 275..299
 /gene="b1028"
 /note="factor Sigma32; predicted +1 start at 1094983"
 389..1375
 promoter
 gene

```
/gene="ycdu"
/note="synonym: b1029"
389. .1375
/gene="ycdu"
/function="orf; Unknown"
/note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCGE
SW: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIPDYLFIRFODKRNLIYIYALIGLIGPYKNAQFPFSEDI
GVSGIALLVNRIFFDLKAYAKCYKRNIDFSPFKKQNHKEIFLTPVAGFSL
LIMSASWGLIQLPSLYALFLSLPLVIFLFRIRTSYKQVIAVAKVKYK
SLTRVLLSVCSITVNLITISPLNNSDFTEGQWLFKSIILALILGVALINLF
FLRSKRYAFLGRFLQIDILFSESNALSTFPAKPLMLRLFILIVEMNITVSVL
ATLVEMRIWPEAYFLCYVPCLIYFYCFRLMNDPMACDMYFRMGHPNK"
1253. .1280
/gene="ycdu"
/note="central position to predicted promoter: -104"
/bound_molecly="Deor predicted site"
1253. .1280
/gene="ycdu"
/note="No predicted promoter"
/bound_molecly="Deor predicted site"
1337. .1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. .1745
/gene="b1030"
1494. .1745
/gene="b1030"
/function="orf; Unknown"
/note="083; This 83 aa ORF is 25 pct identical (7 gaps)
to 81 residues of an approx. 616 aa protein ALBU_RABIT SW:
P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGSTRVOYTKKKPVLSYELFKYGGEGGIDLSRSCGPANH
VHGSSNLSPEVDRSGILLIPVCAIHEKKARTFRALL"
1513. .1926
/gene="ycdu"
/note="synonym: b1031"
1513. .1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS37_YEAST SW:
P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKKSYPFRTSSLIAMAVRGGLTFAPAGSPLTAFVCP
GROLSTPGRFSRCPVQYTKKKVSYELFFKYGGEGGIDLSRSCGPANHCVHGL
SNMISPVPEPSGLIPVCAIYEKARTFRALL"
complement(2111. .2198)
/gene="bex"
/note="synonym: b1032"
complement(2111. .2198)
/gene="bex"
/product="tRNA-Ser"
/note="anticodon: GGA; CG Site No. 17878"
2269. .2296
/function="RNA; tRNA"
2338. .2365
/note="factor Sigma70; predicted +1 start at 1096980"
promoter
```

```
/note="factor Sigma70; predicted +1 start at 1097049"
2355. .2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. .3370
/gene="ycdu"
/note="synonym: b1033"
2393. .3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPA_CORGL
SW: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRIINFAQFSGVSDIIFPHPTPDQWMIABRAIKAIPOAVRAW
KSGNDSADVALVNHPPVEMAGSDIKAVFALGAGVDSILSKLQHPMLNDSVLPFR
LEDYMGEOHMOEYAVSQTLMHFRFPDRIQONSHQPLREYHREDFTIGLGGVL
GSKVAQSLQIWRFPKRSVRKSPGVQSPAGBELSAFLSQCVLITLNPETEV
GIHQQLLEKLPGDAYLTLNLRGVHVEDDLAALDSGVKGMLDVNRBPLPESP
LMQHPRVITTPHVAITRPAEAVYISRTIAQLBKGRVCGQVDRLRGY"
3278. .3308
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1097992"
3383. .3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396. .3423
promoter

Query Match 99.6%; Score 454.4; DB 1; Length 10346;
Best Local Similarity 99.8%; Pred. No. 3.9e-12;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 ATGAACCTTTAAAGTGAAGCAATTCGACAAATCGATTCTCCGATAGCGCTTCGCA 60
8993 ATGAACCTTTAAAGTGAAGCAATTCGACAAATCGATTCTCCGATAGCGCTTCGCA 9052
61 GGTGTTGTTCTCAATACGCGCGCGCGGTAAACCAAGGTGTCGCGTAATATAGCGGC 120
9053 GGTGTTGTTCTCAATACGCGCGCGCGGTAAACCAAGGTGTCGCGTAATATAGCGGC 9112
121 CCAATTCGACGTAACATTACAGTACGTCGTCGTAATCTCGCACTTCGTCGAA 180
9113 CCAATTCGACGTAACATTACAGTACGTCGTCGTAATCTCGCACTTCGTCGAA 9172
181 ACTGATGCCCGTAACCTGACTGATTAACAGATGCGCGCGGTAAATGTCAGAT 240
9173 ACTGATGCCCGTAACCTGACTGATTAACAGATGCGCGCGGTAAATGTCAGAT 9232
241 GTTGCTAGGCGCTAGATGACGCTCAATGATTCGACCAACGTCGTCGTAACAGC 300
9233 GTTGCTAGGCGCTAGATGACGCTCAATGATTCGACCAACGTCGTCGTAACAGC 9292
301 GCTACTCTGATCACTGAGGAAAGCAAAATTCGAAATGAGGTAAACAGTTCGTCGT 360
9293 GCTACTCTGATCACTGAGGAAAGCAAAATTCGAAATGAGGTAAACAGTTCGTCGT 9352
361 GGCACGCGTGTGAGTTCACCAAGTACATCTAATCTCCGTCACAGTACAGTT 420
9353 GGCACGCGTGTGAGTTCACCAAGTACATCTAATCTCCGTCACAGTACAGTT 9412
421 GGCCTTGGTAAACAAGCGACCGCTCATCATGACTAA 456
9413 GGCCTTGGTAAACAAGCGACCGCTCATCATGACTAA 9448

RESULT 5
D90741 15047 bp DNA linear BCT 25-DEC-2002
LOCUS Escherichia coli K12 genomic DNA. (23,7 - 24.0 min).
DEFINITION D90741 AB001340
ACCESSION D90741.1 GI:1651509
VERSION
```

KEYWORDS

Complete and shotgun sequencing; csqG; csqF; csqE; csqD; csqB;

SOURCE

csqG; ycdE; cIs; nov; mdog.

ORGANISM

Escherichia coli K12
Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

REFERENCE

1

Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M., and Horiiuchi, T.,
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T.,
Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

csqG; ycdE; cIs; nov; mdog.

Escherichia coli K12

Escherichia coli K12

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.

1

Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H.,
Nishio, Y., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M., and Horiiuchi, T.,
A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
DNA Res. 3 (3), 137-155 (1996)

2

8905232

Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A., Horiiuchi, T.,
Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K.,
Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M.,
Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H.,
Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T.,
Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.

The systematic sequencing of the Escherichia coli genome in Japan
unpublished

3 (bases 1 to 15047)

Mori, H.

Direct Submission

Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info., 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@gc.aist-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)

Collaboration Information:

Project:

The Japan E.coli genome DNA sequencing project

Group:

The Japan E.coli genome DNA sequencing group

Members: (1995.4-1996.3)

Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Horiiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S.,
Itoh, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S.,
Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K.,
Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K.,
Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N.,
Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.

Headed by:

Name: Takashi Horiiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@gc.aist-nara.ac.jp

URL:
The Japan E. coli genome database
http://sw3.aist-nara.ac.jp

Location/Qualifiers

1. 15047

/organism="Escherichia coli K12"

/mol_type="genomic DNA"

/strain="K12"

/db_xref="taxon:83333"

/map="23.7-24.0 min"

/clone="Kohara clone #231"

/note="Nucleotide position 1097101-1112147 from the
initiation site of Phra (0 min.).-This clone is from
Kohara lambda miniset library."

CDS

CDS

CDS

CDS

CDS

gene

CDS

<1. 276

/note="unnamed protein product; ORF_ID:0230#5"

/codon_start=1

/transl_table=11

/protein_id="BAJ35822.1"

/db_xref="GI:4062605"

/translation="DLFPSSNNALSTFPAPKWLRLFILVIEVMITLVSLATLVE
WRIMEAFELICVPLCIITYFFCYCRFLMNDPMMACDMYFRMCHFNK"

1152. 2090

/note="ORF_ID:0231#1

similar to SwissProt Accession Number P45637"

/codon_start=1

/transl_table=11

/product="Hypothetical 33.0 kd protein in prob-proA
intergenic region"

/protein_id="BAJ35823.1"

/db_xref="GI:4062606"

/translation="MDIIFYPPTQWMEALRAKAPQAVRMKSGDNDSDAYALV
WHPVEMLAGRLKAVFALGAGVDSILSKLOHPMLNPSVPLRIEDTQGEOMBY
AVSQVLMWRFRPDDYRIQONSQSHMOPLREYHREPTGILGAGVLSKVQSLQTFE
PLRCMSRTKSPGVQSPFAGRELSAPISCRVILNLDNTEVETGILINOOLEKLPD
GAYLNLARGVHVEDDILALDSKVYGMALVDVFNNEPPEPSPLMQHPVITTPHY
AATTPAEALVEYISRTIQLERKERVCGQVDRANGY"

2145. 2882

/note="unnamed protein product; ORF_ID:0231#2"

/codon_start=1

/transl_table=11

/protein_id="BAJ35824.1"

/db_xref="GI:4062607"

/translation="WYVDLHMHTVASTAVSTISDYIAQAKGKIKLPAITDHPDM
EDAPPHMFIWRIIMPRVIVGVGILREIENIKNVGEIDDSKMDPSDLIAGHE
PYFAPDKATNTQMTATISGNVHIIISHGNKYEIDVAVVAEAAKQVLEINS
SEIHSRKSPENCSEVAAVADGWAQVLDSDHTAFTWGEFECKIADVDEPPER
ILNYSRRLNLFESRGMAPIAEFADL"

2906. 3460

/note="ORF_ID:0231#3

similar to PIR Accession Number F64035"

/codon_start=1

/transl_table=11

/product="Hypothetical protein HI1543"

/protein_id="BAJ35825.1"

/db_xref="GI:4062608"

/translation="WNEFSILCRVLGSLYRQDPLVPLFTLLRBEKLAANPLEO
DELTUQLQSCDMTQVSADYNALFTGDECAVPPRSAWEGATPAEPAFLISEGML
ADTPAHCHITLLIAGSWLEDDSTEDSEALETLTFSEVILPWCAGFLGVEAHATTPEW
RTVAPLRTDAISAMWDELEEDSEE"

3514. 4053

/note="unnamed protein product; ORF_ID:0231#4"

/codon_start=1

/transl_table=11

/protein_id="BAJ35826.1"

/db_xref="GI:4062609"

/translation="MCSISFYMAAFSAIMRGNNILSIATTGILSGIKWVAVSLGL
SWAGFLGCTAYFACPOGKLGKLAISAATLISGVYMANVITYGALAPHEILGYVTG
IYAFIMCIQAKQOLLSIFVPGTIGACATPAGQGMKVLPSIALGLIFGVMKNSGLW
LAARSAKTAHREOSIKKKA"

complement(4117. 4950)

/gene="csqG"

complement(4117. 4950)

/note="ORF_ID:0231#5

similar to SwissProt Accession Number P52103"

/codon_start=1

/transl_table=11

/product="Curli production assembly/transport component
CsqG precursor"

/protein_id="BAJ35827.1"

/db_xref="GI:4062610"

/translation="MQRFLVAVMLSGCLTAPPKEAAPTMRASQSYDOLTHLPA
PTKIFIVSYNIDQETGQFPPASNFSYAVQSSATMLVYALKDSRFITLEQGLQ
NLNERRKIRIAQENGTVAINNRIPLGISTANIVBSGIIIGSNVKSQGVGARYGQ
IGADTOYDOJAVNIRVNVSTGRTISLSSVSTKITSLEVOAGVFFRIDYORLLBGE
VGYTSNEPMLCLMSAIEGTGVFLINDIGRLMDLQNKAKRQNDIIVKRYHMSVPE

gene S"
complement(4977. .5393)
/gene="csgf"
complement(4977. .5393)
/gene="csgf"
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Csgf precursor."
/protein_id="BAJ35828.1"
/db_xref="GI:4062611"
/translation="MRVHAAVLMLISPLSWAGMTTFOPNPFGNNGAFLNS
AQAQNSYKDPSPYNDPFGIETPSALDFTQALQSQILGGULSNINTGKRGKRVNTDIYI
DIANKDQLQVLDKRTGTSTIQVSGLQNNSTDF"
complement(5418. .5807)
/gene="csgg"
complement(5418. .5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Csgg precursor."
/protein_id="BAJ35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEFLFAAGNLHAEVEVPGLLTDHTVSSIGHDFY
RASFQMSDYDGLTINERPSARWGSWITTVNDVIFQTFLEPKKDFKTVVFAI
IOTEBALNRQIIONLSTGLADHEP"
complement(5812. .6462)
/gene="csgd"
complement(5812. .6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable csgAB operon transcriptional regulatory
protein."
/protein_id="BAJ35830.1"
/db_xref="GI:4062613"
/translation="MFNEVHSIHGHTLLITKSSLOATALLQHKQSIAITGKLNITQ
RSLDDISSGSIITLDMEADKKLIHYODTTSRKNNIKITLLTPEDYPRIDENMP
HINGFYSMEQDERVNGVGLGVEGECYFTQKLASYLLTHSGNRYNSTESALLTHRE
KEINKLRIIGASNNELASLPISNTVXTHLYNLFKLIAYKRTQAVSWANDNLR"
7217. .7672
/gene="csgB"
7217. .7672

gene
CDS

Query Match 99.6%; Score 454.4; DB 1; Length 15047;
Best Local Similarity 99.8%; Pred. No. 4.1e-12;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAGAAGCAATTCAGCAATTCCTCCGTRAGGCTTGCA 60
DB 7713 ATGAAACTTTAAAGTAGAAGCAATTCAGCAATTCCTCCGTRAGGCTTGCA 7772
QY 61 GGTGTTGTTCTCACTAGACGGCGCGCGGTAAACACGCTGGTGGCGTAATATAGCGGC 120
DB 7773 GGTGTTGTTCTCACTAGACGGCGCGCGGTAAACACGCTGGTGGCGTAATATAGCGGC 7832
QY 121 CCAAAATTCGAGCTGAACATTACAGTAGACGCTGGCGGTAACTGTGCACTTGCTTGCA 180
DB 7833 CCAAAATTCGAGCTGAACATTACAGTAGACGCTGGCGGTAACTGTGCACTTGCTTGCA 7892
QY 181 ACTGATGCGCGTACTGACTGACTATTAACCAAGCATGGCGCGGTAAATGTCACAT 240
DB 7893 ACTGATGCGCGTACTGACTGACTATTAACCAAGCATGGCGCGGTAAATGTCACAT 7952
QY 241 GTTGGTCAGGCGTACATGACATCAATCATCTGACCAACGTCGCTCGTAAACAGC 300

DB 7953 GTTGGTCAGGCGTACATGACATCAATCATCTGACCAACGTCGCTCGTAAACAGC 8012
QY 301 GCTACTTTGATCACTGGAACGGCAAAATTTCTGAATGACGTTAAACAGTTGGTGGT 360
DB 8013 GCTACTTTGATCACTGGAACGGCAAAATTTCTGAATGACGTTAAACAGTTGGTGGT 8072
QY 361 GGCAACGGTGCCTGACGTTGACAGCTGACATCTAATCTCTCCGTCGAACGTCAGGTT 420
DB 8073 GGCAACGGTGCCTGACGTTGACAGCTGACATCTAATCTCTCCGTCGAACGTCAGGTT 8132
QY 421 GGCTTTGGTAAACAACGACCGCTCATCATGACTTAA 456
DB 8133 GGCTTTGGTAAACAACGACCGCTCATCATGACTTAA 8168

RESULT 6
AF275733 1711 bp DNA linear BCT 15-MAY-2001
LOCUS AF275733
DEFINITION Escherichia coli strain 43895 Red Variant CsgB protein (csgB) and
AF275733 CsgA protein (csgA) genes, complete cds.
ACCESSION AF275733
VERSION AF275733.1 GI:14039399
KEYWORDS
SOURCE
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 (bases 1 to 1711)
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Mutations in the csgD promoter associated with variations in curli
expression in certain strains of Escherichia coli O157:H7
Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
JOURNAL MEDLINE
21218556
REFERENCES
11319125
2 (bases 1 to 1711)
Ulrich, G.A., Keen, J.E. and Elder, R.O.
Direct Submission
JOURNAL
Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
Research Center, State Spur 18D, Clay Center, NE 68933, USA
FEATURES
source
1. .1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757. .1212
/gene="csgB"
757. .1212
/gene="csgB"
/codon_start=1
/transl_table=1
/product="CsgB protein"
/protein_id="AAK53211.1"
/db_xref="GI:14039400"
/translation="MKRLKPMULTTIGARCIATAAAGYDLSNRYNAVELSSRPN
OATITGAGTNSAQLRQGSKLAIVAQESSRRAKIDGTGDLNLYITDQASANDA
SISQAGANTAMIIQKSGNKNANTYGTOKTALIVOROSMAIRVOR"
1253. .1711
/gene="csgA"
1253. .1711
/gene="csgA"
/codon_start=1
/transl_table=1
/product="CsgA protein"
/protein_id="AAK53212.1"
/db_xref="GI:14039401"
/translation="MKRLKVAATTAIYPSGSAIAGVYPOYGGGNGHGGGNSGPN
ELNTYQGGNSALALQADRNSTLTITQGGGAGAVGGSDSSIDLTQRFGNNA
TLQWNGKDSHMTYKQFGGNGAADVDTANSTVNTVQFGNNATAHQY"

gene
CDS

ORIGIN
Query Match 86.6%; Score 395; DB 1; Length 1711;

Best Local Similarity 92.8%; Pred. No. 6,3e-104;
Matches 426; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

```
QY 1 ATGAACCTTTTAAAGTGAAGCAATTGCGCATGTGATCTCCGAGCGCTCTGGCA 60
Db 1253 ATGAACCTTTTAAAGTGAAGCAATTGCGCATGTGATCTCCGAGCGCTCTGGCA 60
QY 61 GGTGTGTCTCTCACTGATCGCGCGC---GGCGGTAAACACCGGGGTGGGTGAATATAGC 117
Db 1313 GGTGTGTCTCTCACTGATCGCGCGC---GGCGGTAAACACCGGGGTGGGTGAATATAGC 117
QY 118 GGCCCAATTTCTAGTGAACATTTACAGTACCGTGGCGGTAACTCTGCACTTGGCTGTG 1372
Db 1373 GGCCCAATTTCTAGTGAACATTTACAGTACCGTGGCGGTAACTCTGCACTTGGCTGTG 1372
QY 178 CAATGTATGCCCTGTAACCTGACTGCTATTAACCGAGTACGGCGGTGAATAGTGA 1432
Db 1433 CAATGTATGCCCTGTAACCTGACTGCTATTAACCGAGTACGGCGGTGAATAGTGA 1432
QY 238 GATGTTGGTCAGGGCTCAGATGACAGCTCAATGATGATGACCGAAGTGGCTTGGTAC 1492
Db 1493 GATGTTGGTCAGGGCTCAGATGACAGCTCAATGATGATGACCGAAGTGGCTTGGTAC 1492
QY 298 AGCGTACTCTTGTATGATGAGTGAACCGGAAATTTCTGAATGACCGGTAAACAGTGGT 1552
Db 1553 AGCGTACTCTTGTATGATGAGTGAACCGGAAATTTCTGAATGACCGGTAAACAGTGGT 1552
QY 358 GGTGGCAACGGTGTCTGAGTGAACCGAGCTGATGATGATGATGATGATGATGATGATG 1612
Db 1613 GGTGGCAACGGTGTCTGAGTGAACCGAGCTGATGATGATGATGATGATGATGATGATG 1612
QY 418 GTTGGCTTTGTTGAACACCGACCGAGCTGATGATGATGATGATGATGATGATGATG 1672
Db 1673 GTTGGCTTTGTTGAACACCGACCGAGCTGATGATGATGATGATGATGATGATGATG 1672
```

RESULT 7
LOCUS AE005315 10190 bp DNA linear BCT 21-MAR-2001
DEFINITION Escherichia coli O157:H7 EDL933 genome, config 1 of 3, section 139
ACCESSION AE005315 AE005174
VERSION AE005315.1 GI:12514572
KEYWORDS
SOURCE
ORGANISM Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
JOURNAL MEDLINE
PUBMED 21074935
FBI 11206551
TITLE 2 (bases 1 to 10190)
AUTHORS Perna, N.T., Plunkett, G. III, Burland, V., Mau, B., Glasner, J.D.,
Rose, D.J., Mayhew, G.F., Evans, P.S., Gregor, J., Kirkpatrick, H.A.,
Postfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L.,
Grothbeck, E.J., Davis, N.W., Lim, A., Dimalanta, E., Potamousis, K.,
Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,
Welch, R.A. and Blattner, F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
JOURNAL
TITLE
SOURCE
Location/Qualifiers
1..10190
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"

```
gene  
/strain="EDL933"  
/set_type="O157:H7"  
/db_xref="taxon:155864"  
/note="enterohaemorrhagic"  
617..1072  
/gene="cs98"  
/note="synonym: Z1675"  
617..1072  
/gene="cs98"  
/function="structural component; Cell exterior  
constituents: Surface structures"  
/note="Residues 1 to 151 of 151 are 100.00 pct identical  
to residues 1 to 151 of 151 from Escherichia coli K-12  
Strain MG1655: B1041"  
/codon_start=1  
/transl_table=11  
/product="minor curlin subunit precursor, similar to CsgA"  
/protein_id="AAG55787.1"  
/db_xref="GI:12514573"  
/translation="MKNLKPLMVLITLGPGLAAGYDANSEINPAVNELSKSFN  
STSGAVGNTAMITQKSGNKRANITQYGTQKTAIVQROSAIRVQR"  
1113..1571  
/gene="cs98"  
/note="synonym: Z1676"  
1113..1571  
/gene="cs98"  
/function="regulator; Cell exterior constituents: Surface  
structures"  
/note="Residues 1 to 152 of 152 are 96.71 pct identical to  
residues 1 to 151 of 151 from Escherichia coli K-12 Strain  
MG1655: B1042"  
/codon_start=1  
/transl_table=11  
/product="curlin major subunit, coiled surface structures;  
cryptic"  
/protein_id="AAG55788.1"  
/db_xref="GI:12514574"  
/translation="MKNLKPLMVLITLGPGLAAGYDANSEINPAVNELSKSFN  
STSGAVGNTAMITQKSGNKRANITQYGTQKTAIVQROSAIRVQR"  
1530..1962  
/gene="cs98"  
/note="synonym: Z1677"  
1530..1962  
/gene="cs98"  
/function="putative structure; Not classified"  
/note="Residues 1 to 110 of 110 are 98.18 pct identical to  
residues 1 to 110 of 110 from Escherichia coli K-12 Strain  
MG1655: B1043"  
/codon_start=1  
/transl_table=11  
/product="putative curli production protein"  
/protein_id="AAG55789.1"  
/db_xref="GI:12514575"  
/translation="MNALILAAASQTTNTTQGGDMYTIPEVLTGSCRYOIL  
WPSSEKKS"  
2083..2394  
/gene="cs98"  
/note="synonym: Z1678"  
2083..2394  
/gene="cs98"  
/function="orf; Unknown function"  
/note="Residues 1 to 103 of 103 are 98.05 pct identical to  
residues 1 to 103 of 103 from Escherichia coli K-12 Strain  
MG1655: B1044"  
/codon_start=1  
/transl_table=11  
/product="orf, hypothetical protein"  
/protein_id="AAG55790.1"  
/db_xref="GI:12514576"  
/translation="MRPFLNLSIMLGSLLFPPFIAGSTAGQGIHFYGOIVPACDV  
STOSSPEVNCPPONGSVPGKTYSSKALMSGNVKAQJASVYQYLDKOKLAVNIEY
```


TITLE	Complete nucleotide sequence of the prophage VT2-Sakai carrying the verotoxin 2 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 derived from the Sakai outbreak
JOURNAL	Genes Genet. Syst. 74 (5), 227-239 (1999)
MEDLINE	20198780
PUBMED	10734605
REFERENCE	2
AUTHORS	Ohnishi, M., Murata, T., Nakayama, K., Kuhara, S., Hattori, M., Kurokawa, K., Yasunaga, T., Yokoyama, K., Makino, K., Shinagawa, H. and Hayashi, T.
TITLE	Comparative analysis of the whole set of rRNA operons between an enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain and an <i>Escherichia coli</i> K-12 strain MG1655
JOURNAL	Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE	20557356
PUBMED	11108008
REFERENCE	3
AUTHORS	Yokoyama, K., Makino, K., Kubota, Y., Watanabe, M., Kimura, S., Yutsudo, C.H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T., Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and Shinagawa, H.
TITLE	Complete nucleotide sequence of the prophage VT1-Sakai carrying the Shiga toxin 1 genes of the enterohemorrhagic <i>Escherichia coli</i> O157:H7 strain derived from the Sakai outbreak
JOURNAL	Gene 258 (1-2), 127-139 (2000)
MEDLINE	20564182
PUBMED	111101050
REFERENCE	4
AUTHORS	Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T., Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C., Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and Shinagawa, H.
TITLE	Complete genome sequence of enterohemorrhagic <i>Escherichia coli</i> O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL	DNA Res. 8 (1), 11-22 (2001)
MEDLINE	21156231
PUBMED	11258796
REFERENCE	5
AUTHORS	(bases 1 to 327773) Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and Hayashi, T.
TITLE	Direct Submission
JOURNAL	Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871, Japan (E-mail: ken@gen-info.osaka-u.ac.jp), URL: http://www.gen-info.osaka-u.ac.jp/ , Tel: 81-6-6879-8365, Fax: 81-6-6879-2047)
MEDLINE	
PUBMED	
REFERENCE	
AUTHORS	
COMMENT	genome project.
FEATURES	location/Qualifiers
source	1. 327773
gene	/organism="Escherichia coli O157:H7"
CDS	/mol_type="Genomic DNA"
	/strain="O157:H7"
	/sub_strain="RIMD 0500952"
	/db_xref="taxon:83334"
	150..2258
	/gene="ECs1032"
	150..2258
	/note="similar to YCBY ECOLI g11787181 percent identity 99 in 702 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="putative oxidoreductase"
	/protein_id="BAB34455.1"
	/db_xref="GI:13360492"
	/translation="MNSLPASTKRGLEELKTELNLGAECQVQVGVHFKDPTRLV YGLMWSRLASRIIMPLGECIKYSDLDLVGLQVAINWTFENPGATFVHFGSLNDIT RNSQYAMKVDIAIVDAFTRKMLPENVDKADDIRVWVWLHETASIALDLSGGH LRSYDRAGIADIKETLAAIIVRSQMGPGTLPDPCGSGTLLIEAALDADRSGL HRRWFGSGMAQDEAIQOEVAEAKOETARKGLAYSSHPFGSGSDAIVORATNKR LNGIGRLITFEVDVAQLTNPPLPKBYGTIVLSNPYGRLDSEPLALINSLGRIMK
	NOFGWNLSLFSASDILSLCQLRADKQYAKNGPLDQKQNVHASTPDSKPAMWA EDYANRLKLNKKEKWAROGIECYLYADAPENVVNAVDRADWVVEYAPRPTT DAHKAQRLFDPIIATISVLGIAPNKLITRRKQYKQYQGLGEGLEVEYETNA HLAVNLIDYLDGTGLFDRIARMLGMSGKXPLNLFSTYSATVHAGIGGARSFTT VDMRTYLEAERMLPLNGLTGRHRLIQDCIAMEANEQFDLIPIPTPSNKR MEDAPDIPORHIALMKDLKRLNRAAGTITMSNNKRGFRMDLGLAKGAKORITQKT LSQPPARRQIHNCLITAA"
	2270..4177
	/gene="ECs1033"
	2270..4177
	/note="ECs1033"
	/note="similar to UUP ECOLI g11787182 percent identity 99 in 635 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="putative ATP-binding component of a transport system"
	/protein_id="BAB34456.1"
	/db_xref="GI:13360493"
	/translation="MSLISMGAMLSFSDAPLIDNAELIHEDNERVLCVGNAGKST LAKTLIRGGLDGRITLYQDILVARLQDDPPRNVEGVYFVAGIEEOALEYKRYH DLSRLVNDPSSKNNELAKVOEDHNLMOENRINEVIAQGLDNPVLSISG WLRKALGRALVSNPRLIDEPYHNLIDETIDLEGLKPTNGTILFISDPSIRN MATRIVDLRGKLVTPGNYDQYLKEKEALRVEELONAEFPTKLAQGEVATRGITKA RRTNREGVYALKAMRERGERREVMGTAKQVEASRSGLIVBMEVCTVQVQKLA VDFSAQVLRGDKIALIGPNCGKTTLLKMLGOLADSGIHVGTLEKLYAFDQKRA ELDPKTMNDINABKQVYMNVRKRVHVLQDLPFKPKAMTPVRLSGSENNRLI IARLIPKSNLILDEPTNDLVEITLLELILSYQCTVLPKAMTPVRLSGSENNRLI TEBGLQGRVYGGVHDARGOQDQVVALQKAVKNESPAKLEYARSSTKSYTL QNELQQLPQLLEDAKLEALQTVADASFPSQHEQKVLADMAAEOELBEAFER MYLEALNKG"
	4307..5560
	/gene="ECs1034"
	4307..5560
	/note="ECs1034"
	/note="similar to PQA ECOLI g11787183 percent identity 99 in 417 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="paratub-inducible protein A"
	/protein_id="BAB34457.1"
	/db_xref="GI:13360494"
	/translation="WCEHHHAHHLCSQCDMLVALPRLEHGGKACPRGCTTLTVAM DAPRQPTAVALALFMLLSNLPFNMMVAGVTSITLLEIRGLVPSFSDVASIGTF FLFVQVLPARFCLITLLVNRALPVLKQOLARVLFOLKTMGAETFLACVYSFV KLAAYGSIQVSSFLPWCFCVQLAPFOCDVRMLMDIADAPRLROPLKGVGIR QGLRSQSCCTAILPADPVCPCGCTGVRRRSLQMTLALVTSINLTPANLIPGM VTDLGSKMSTIILAGVILMSRGSYPVAVIFLSTINPTLKMIAVLCPWDAGHG KRDEHMLIYEVVEPVGRKSNIDVFIVLSLVVMGLMSIYPMAGLMALVAVIM TMSAMTFDRRLWDRQSEHES"
	5565..7205
	/gene="ECs1035"
	5565..7205
	/note="ECs1035"
	/note="similar to POIB ECOLI g11787184 percent identity 99 in 546 aa (Conserved in E. coli K-12)"
	/codon_start=1
	/evidence=not_experimental
	/transl_table=11
	/product="paratub-inducible protein B"
	/protein_id="BAB34458.1"
	/db_xref="GI:13360495"
	/translation="MESNNGRAKIOKYNMSPWIFPTVALIGAWLFYHSHQGE VTLITANARGIRGCKTITKSRSDVGVGVSTLADDLTHVILKARLNSGEMKLAHOT VDSKKAQGLSGDVLFRGVRVGSVESTDTQKRNISYOLFNAFYDRLVTSNVP WDSQSIADVLISAGRVNMGSLITLLSGVAFVDEGLDQCPAPAKTAFTLYDDOS IDDSIYTHLIDLMFKDSVGLQGLGAVVEGRGRLGTVSVVPPAPAMNOTPDDYR IPVLIRIPEERKMLQGNADVHHLGELIKRGIRGSLKTNLVTGALLYDLDFYVNT PATIGIRENGQIILPTVSGGLAQIQRMALDKLNKZLPNPMIETATSTLESQRT MKNLQTLDSNNKLIASQSMQQLPTDMQSTLREKMQGQPPGSAAYNKNVADMQRL

gene
D0VLEI0PVLKTLNKSNAIVFAKDKDEPKAKK"
7217..7765
/gene="EC01036"
7217..7765
/gene="EC01036"
/note="similar to YMA_ECOLI gi|1787185 percent identity
98 in 182 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="hypothetical protein"
/protein_id="BAB34459.1"
/db_xref="GI:13360496"
/translation="MTALMLAGSSGEINKYOLPVVOSGOSTASOGNRLIMVE
OYAVPDVAGNVYVDSYVYANNMLASPLDOQLRNLVNLSTOLGKMYVASO
PLGSAQDTLNTVTEFNKRYGKTVSGEMLLNHQGLIKRPFRLGEGVQTDGIDEMV
KVLGWSQEAASIAQETIKRLP"
8021..8188
/gene="EC01037"
8021..8188
/gene="EC01037"
/note="identical to RMF_ECOLI gi|1787186 (Conserved in
E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=1
/product="ribosome modulation factor"
/protein_id="BAB34460.1"
/db_xref="GI:13360497"
/translation="MKRKRDRLERARQRYOAGIAGRSKEMCPYOTLNQRSLWGM
REMAADRVVMA"
complement(8258..8776)
/gene="EC01038"
complement(8258..8776)
/gene="EC01038"
/note="similar to FABA_ECOLI gi|1787187 percent identity
100 in 172 aa (Conserved in E.coli K-12)"
/codon_start=1
/evidence=not experimental
/transl_table=1

Query Match 86.6%; Score 395; DB 1; Length 327773;
Best Local Similarity 92.8%; Pred. No. 1.3e-103;
Matches 426; Conservative 0; Mismatches 30; Indels 3; Gaps 1;

gene
CDS
1 ATGAACTTTTAAAGTGAAGCAATTCAGCAATCGATTCTCGGTAAGCGCTCTGCA 60
326421 ATGAACTTTTAAAGTGAAGCAATTCAGCAATCGATTCTCGGTAAGCGCTCTGCA 326480
61 GGTGTGTTCTCTGAGTGGGGG---GGCGGTAAACAGCGGTGGGCGGTAAATATG 117
326481 GGTGTGTTCTCTGAGTGGGGG---GGCGGTAAACAGCGGTGGGCGGTAAATATG 326540
118 GGCCCAATTTGAGCTGAACATTTACAGTACGCTGGGTAACTCTGCACTTGCTCTG 177
326541 GGCCCAATTTGAGCTGAACATTTACAGTACGCTGGGTAACTCTGCACTTGCTCTG 326600
178 CAAACTGATGCCCGTAATCTGACTTACATTAACCAAGATGCGCGCGGTAAATGCTCA 237
326601 CAACTGATGCCCGTAATCTGACTTACATTAACCAAGATGCGCGCGGTAAATGCTCA 326660
238 GATGTGGTCAAGGGCTGAGTGAAGCTCAATTCATGACCCCAACGTTGGCTGATAC 297
326661 GATGTGGTCAAGGGCTGAGTGAAGCTCAATTCATGACCCCAACGTTGGCTGATAC 326720
298 AGCGCTACTCTTGAATCACTGGAACCGCAAAATTTCTGAATGACGCTTAAACAGTTCCGT 357
326721 AGCGCTACTCTTGAATCACTGGAACCGCAAAATTTCTGAATGACGCTTAAACAGTTCCGT 326780
358 GGTGGCAACGGTGTCTGACGACGACGCTGATTAATCTCTCGGTAAAGTGAATCAG 417
326781 GGTGGCAACGGTGTCTGACGACGACGCTGATTAATCTCTCGGTAAAGTGAATCAG 326840

Qy 418 GTTGCTTTGGTAAACGACCGCTCATCACTAA 456
Db 326841 GTTGCTTTGGTAAACGACCGCTCATCACTAA 326879

RESULT 9
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
PUBMED
REFERENCE
AUTHORS
FEATURES
SOURCE
gene
CDS

AE016759 306358 bp DNA linear BCT 24-DEC-2002
Escherichia coli CFT073 section 5 of 18 of the complete genome.
AE016759 AE014075
AE016759.1 GI:26107527
Escherichia coli CFT073
Escherichia coli CFT073
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 306358)
Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
Raeko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J.,
Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blattner, F.R.
Extensive Mosaic Structure Revealed by the Complete Genome Sequence
of Uropathogenic Escherichia coli
Proc. Natl. Acad. Sci. U.S.A. 99 (26), 17020-17024 (2002)
2 (bases 1 to 306358)
Welch, R.A., Burland, V., Plunkett, G.D. III, Redford, P., Roesch, P.,
Raeko, D.A., Buckles, E.L., Liou, S.-R., Boutin, A., Hackett, J.,
Stroud, D., Mayhew, G.F., Rose, D.J., Zhou, S., Schwartz, D.C.,
Perna, N.T., Mobley, H.L.T., Donnenberg, M.S. and Blattner, F.R.
Direct Submission
Submitted (20-JUN-2002) Genetics Laboratory, University of
Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
Location/Qualifiers
1..306358
/organism="Escherichia coli CFT073"
/mol_type="genomic DNA"
/strain="CFT073"
/db_xref="taxon:199310"
complement(83..3820)
/gene="iroC"
/locus_tag="c1253"
complement(83..3820)
/gene="iroC"
/locus_tag="c1253"
/note="Residues 27 to 1231 of 1245 are 80.16 pct identical
to residues 1 to 1209 of 1218 from GenPept.129 :
>emb|CAD05883.1| (A1627276) putative ABC transporter
protein [Salmonella enterica subsp. enterica serovar
Typhii]"
/codon_start=1
/transl_table=1
/product="ATP binding cassette (ABC) transporter homolog"
/protein_id="AA097927.1"
/db_xref="GI:26107528"
/translation="MPDSYLAIRNDPASGVYTPDGGGLMPANHTPPASQWIVLA
RVCKERKLSVTVVAVSTILLALPLTRQAVNDALAGNPRALMLACGLLIAR
FDLIGNVRRGAVAMLSLWQHTLRGVPSIQLDVAGDPLATGVYISRTNSDLOQ
VHTLQMPVPLAVFYIAGIVMLMMSAPMTLIVCVCLAITLRARRVFAQT
GMSADLANTEHIREVLAQISVKSQVAAERETHMIDRSRQIVRRVIGVISQAMP
GATMLALPVIGQIVILCYGMSVNHGIDIGTFAPAFSLPFLMLPVRVLSFLVIL
RTQASVERVALIDTRSDMEDGTESINSQGVKTYANTSEGRONLRDRLERRRVGI
VFEPAFLFAGTVENLAVYGHQATADDIRPAAAGSDFINALPKPDSILTERGN
LSGQQRRIALRALITAPDVLIDDTTSAVDVATEAINTALGRVADBEHMLVIR
RSTLTQASVVVLDKRMVDTGTPALERCAFRAFLMGTGSDPLFATSHNSHMLP
ABEPATDDVDVDTGKGFVAMTRVPEVAVOOALAGKRTSLKLPVAMPEVIALLI
ALPSAAAGVTLITLQHGIDSGVAAGMSTIGCALLALCIYVGMCSVSLQTVARA
AESYQNSVRLRSRGNRLRLGIPMERKADSRILRMVTDVDSIARFLONGLAGATSIV
TMFPAITMTWLDITLALISAVPVALATMTYRISTPAVAPLFEICGASVAAVL
KVGSLRVQSHGQDELEGARILRAISFRAITRVAVVPEPLTFCEASVAAVL
LVGASQVAAQEMTAVGLAAFPFLIGQYGVQVQVSGIVDMQOATAGAKIDELATE
GTENLGSSTVLPVGTALHLDVTFPSYDPSHSBPANLKLTLITPEGVAVAVGSRGAGS

Qy	118	GGCCCAAAATTCAGCTGAACATTACAGTAGTCGGTGGCGGTAACCTGACCTTGCTCTG	177
Db	45389	GGTCCGAATTGAGAGCTGAAATTATTACAGTAGTCGGTGGCGGTAACCTGCTTGCTCTAG	45448
Qy	178	CAAACTGATGCCCGTAATCTTGACTTGACTATTATCCAGCATGGCGGCGTAATGTCGA	237
Db	45449	CAAGTCAGCGCCCGTAACCTGATCTGACCAATTACCAGCAGCGGCGGTAATGGCCCA	45508
Qy	238	GATGTGGTCAAGGGCTCAGATGATGACAGCTCAATCATGATGACCAAGTGGCTGGTAC	297
Db	45509	GATGGGGCCCAAGGTTCTGATGACAGCTCAATCATGATGACAGCTGGTGGTGGCAAC	45568
Qy	298	AGCGCTACTCTTTGATCAGTGGAAACGCAAAAATTCGAATGACGCGTTAAAACAGTTCCGT	357
Db	45569	AGCGCTACTCTTTGATCAGTGGAAATGTAAGATTCTACTATGACTGTTAAACAGTTCCGT	45628
Qy	358	GGTGCAACGGGTGCGAGCTGACGACGACGCTGATTAATCTCTCGGTCAAGTACTGAC	417
Db	45629	GGCGGTAACGGGTGCGAGCTGATGACGACGCTGATTAATCTCTCGGTCAAGTACTGAC	45688
Qy	418	GTGGCTTTGGTGAACAACGACGACGCTCATGACTACTAA	456
Db	45689	GTGGCTTTGGTGAACAACGACGACGCTCATGACTACTAA	45727
RESULT 10			
LOCUS	AE015131	10370 bp	DNA
DEFINITION	Shigella flexneri 2a str. 301	section 94 of 412	of the complete genome.
ACCESSION	AE015131	AE005674	
VERSION	AE015131.1	GI:24051313	
KEYWORDS			
ORGANISM	Shigella flexneri 2a str. 301		
SOURCE	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.		
REFERENCE	1 (bases 1 to 10370)		
AUTHORS	Jin, Q., Yuan, Z.H., Xu, J.G., Wang, Y., Shen, Y., Lu, W.C., Wang, J.H., Liu, H., Yang, J., Yang, F., Qu, D., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., He, B.K., Chen, R.S., Ma, D.L., Qiang, B.Q., Wen, Y.M., Hou, Y.D. and Yu, J.		
TITLE	Genome sequence of Shigella flexneri 2a: insights into pathogenicity through comparison with genomes of Escherichia coli K12 and O157		
JOURNAL	Nucleic Acids Res.	30 (20), 4432-4441	(2002)
PUBMED	12384590		
REFERENCE	2 (bases 1 to 10370)		
AUTHORS	Jin, Q., Shen, Y., Wang, J.H., Liu, H., Yang, J., Yang, F., Zhang, X.B., Zhang, J.Y., Yang, G.W., Wu, H.T., Dong, J., Sun, L.L., Xue, Y., Zhao, A.L., Gao, Y.S., Zhu, J.P., Chen, S.X., Yao, Z.J., Wang, Y., Lu, W.C., Qiang, B.Q., Wen, Y.M. and Hou, Y.D.		
TITLE	Submitted (08-MAY-2001) Microbial Genome Center of Chinese Ministry of Public Health, 100 Yingxin Jie, Xuanwu Qu, Beijing 100052, P.R. China		
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
PUBMED			
REFERENCE			
AUTHORS			

```

/insertion="1"
/insert_table="1"
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="gi:24051317"
/translation="MCELDIHDSLVQPCPELHKRLNSLTACNALDCKLITTEIR
GNLPTKATKTHNKKRIDRLGNHSLKEELAYRMHSAFISGNTMPTLVADMSDIFR
BOKRLMTRASVALHSGSVLYEAFAPLSQCSKKHDFDLASILPSNTPLTIVASD
DAGEFVPMYKVEELGKQWMLSRVAFQSYVDIGAEWKKRIKSLDHMSHSSHLKELG
LTKPSNLSQCLITLSSKSKRNQKRSRTHCHHPKSKITSASAKPMNSLNRLVELEIR
QKPLIVNISAKRMQIEETFPDLKSPVAGLGRHSPSTSSSEFIDMLLTALMLQTCWL
AGVAGKQGMGDKHQQANTVNRNVLSLTGLGMEVLHSGVITTRDESLVATLTLTQVL
FTTHGVGIKK"
Complement (5995..6384)

```

```

//locus tag="SF1033"
//note="Residues 1 to 129 of 129 aa are 98 pct identical to
residues 1 to 129 of a 129 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_281173.1"
Codon start=-1

```

```

//locus_tag="SF1034"
//note="Residues 11 to 226 of 226 are 100 pct identical to
residues 1 to 216 of a 216 aa protein from Escherichia
coli O157:H7 EDL93 ref: NP_287174.1"
//codon_start=1

```

gene	7766. .8248	/gene="csGB"
		/locus_tag="SF1035"
CDS	7766. .8248	

```

/locus_tag="SF1035"
/note="Residues 10 to 160 of 160 are 99 pct identical to
residues 1 to 151 of a 151 aa protein from Escherichia

```

QY	105	CGGTAATATATAGCGGCCCAATTCTGAGCTGAACATTACAGTACGGTGGCGGTAACTTC	164
Db	9538	CAGTATATATATATAGCGGCCCAAAATTCAGAGCTGAACTTTACAGTACCGGTGGCGTAACTTC	959
QY	165	TGCACTTGCTCTGCAAACTGATGACCCTGTAATCTGACATCTGATATTTTCCAGATGCGG	224
Db	9598	TGACCTTGCTCTGCAAACTGATGACCCTGTAATCTGACATCTGATATTTTCCAGATGCGG	965
QY	225	CGGTAATGAGTGCAGATGCTTGGTCACGGGCTCAGATGACAGCTCAATGATCTGACCCAAAG	284
Db	9658	TGGTATATGAGTGCAGATGCTTGGTCACGGGCTCAGATGACAGCTCAATGATCTGACCCAAAG	971
QY	285	TGGCTTCGGTAAACACCGTACTCTTTGATCAATGGAACGGCAAAATTTGCAATGACGT	344
Db	9718	TGGCTTCGGTAAACACCGCGCACTCTTTGATCAATGGAACGGCAAAATTTGCAATGACGT	9777
QY	345	TAAACAGTTCCGTGGTGGCAACCGGTGCTGACGTTGACCAAGATCTGAATCTCTCCGT	404
Db	9778	TAAACAGTTCCGTGGTGGCAACCGGTGCTGACGTTGACCAAGATCTGAATCTCTCCGT	9837
QY	405	CACCGTACATGAGTTGGCTTTGGTAAACAACGGACCGGTATCAGTACTAA	456
Db	9838	CACCGTACATGAGTTGGCTTTGGTAAACAACCGACCGGTATCAGTACTAA	9889

	RESULT	11			
	LOCUS	AE016981			
	DEFINITION	AE016981 Shigella flexneri 2a str.	292S04 bp DNA	linear BCT 22-APR-2003	
	ACCESSION	Genome AE016981, AE014073			
	VERSION	AE016981.1 GI:30040616			
	KEYWORDS				
SOURCE ORGANISM					
		Shigella flexneri 2a str.	2457T		
		Shigella flexneri 2a str.	2457T		
		Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
		Enterobacteriaceae; Shigella.			
REFERENCE		1 (bases 1 to 292504)			

REFERENCE
AUTHORS
1 (bases 1 to 292504)
Wei, J., Goldberg, M. B., Burland V. von-

TITLE Complete Genome Sequence and Comparative Genomics of *Shigella flexneri* and Discher, F.R.
JOURNAL *flexneri* Serotype 2a Strain 24577
PIBMED Infect. Immun. 71 (5), 2775-2786 (2003)
12704152

REFERENCE AUTHORS

TITLE Direct Submission
JOURNAL Submitted (13-JUN-2002) Genetics Laboratory, University of Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA
FEATURES Location/Qualifiers
source 1..292504

```

/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"

```

```
gene /serotype="2a"
/db_xref="taxon:198215"
/complement(223..441)
/gene="infA"
/locus_tag="S0884"
/complement(223..441)
/gene="infA"
/locus_tag="S0884"
/function="factor; Proteins - translation and modification"
/note="residues 1 to 72 of 72 are 100.00 pct identical to residues 1 to 72 of 72 from Escherichia coli K-12 : B0884"
/codon_start=1
/transl_table=11
/product="protein chain initiation factor IF-1"
/protein_id="Aap16348.1"
/db_xref="GI:30040617"
/translation="MAKEDNIEMQGTVELTPNTWFRVLENGHVTAHISGKMKRY
RILTGDRVTELPYDSKGRIVRSR"
/complement(726..1430)
/gene="aat"
/locus_tag="S0885"
/complement(726..1430)
/gene="aat"
/locus_tag="S0885"
/function="enzyme; Aminoacyl tRNA synthetases, tRNA modification"
/note="residues 1 to 234 of 234 are 100.00 pct identical to residues 1 to 234 of 234 from Escherichia coli K-12 : B0885"
/codon_start=1
/transl_table=11
/product="leucyl, phenylalanyl-tRNA-protein transferase"
/protein_id="Aap16349.1"
/db_xref="GI:30040618"
/translation="MRVLQSRHSIAPSPPEGALKEPNGLALGDSLSPARLMAVOR
GIPWFSGDDILMWSPPRAVLNPSLSHISGKMRPKRSPRYTMVAAGOVYEGC
ASDRBECTIRGVVEAYRHLDELGHASISVEMREDELVGCMYGAQOTLTCGSMPS
RMEANSKTALVFCBEFTGHGSKLIDCVLNDHTASLACSEIPRDYLYNLQMKELGR
LPNNFWPRCLFSPE"
/complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/complement(1472..3193)
/gene="cydC"
/locus_tag="S0886"
/function="transport; Transport of small molecules: Other"
/note="residues 1 to 573 of 573 are 97.03 pct identical to residues 1 to 573 of 573 from Escherichia coli K-12 : B0886"
/codon_start=1
/transl_table=11
/product="ATP-binding component of cytochrome-related transport"
/protein_id="Aap16350.1"
/db_xref="GI:30040619"
/translation="MRALLPYLALYKRKMLSGIVLAIVTLASIGLTLSCMFSL
ASAVAGVGLYSFNMLPAAGVGAITRTAGRFERLVSDATFRVQLHRTYFSK
LLPLSPAGLARYQOGLNRYVADVDLIDHLYLRIYSPVGAFAVIMVITGLFIDK
TLATLGSVMLTLFLMPPLFYRAGKSTGONLTHLRGOOLTAMLQGAELTFGA
SDRYTOLENTIOMLEAOROSSETLALSOIMLLIGALAVILMMASGGVGNAP
GALIALPYFCALAPREALAPYTGAFQHLGQVITASAVRSIDTDQKEVTPPDYTRVA
DRVSTLTDVQPTTPESQNALKGISLVNAGEHAIIGRTGCGSTLQULTRAMPD
QQGSETLNDSPIASINENALROTISVVPQARFENAIIRDLMLASPOSSDBALSEIL
RVRGIEKLBDAGLNSWLEGGRQLSGGLRRLAIARTLHDAPLVLDDETEGDAT
TESQIELLAKMREKTVLWTHRLGRISRQOIIIMDNGQIIEGGTHAEILAGQRY
YQPKQGL"
/complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"
/complement(3194..4960)
/gene="cydD"
/locus_tag="S0887"

gene /function="transport; Not classified"
/note="residues 1 to 588 of 588 are 92.85 pct identical to residues 1 to 588 of 588 from Escherichia coli K-12 : B0887"
/codon_start=1
/transl_table=11
/product="ATP-binding component of cytochrome-related transport"
/protein_id="Aap16351.1"
/db_xref="GI:30040620"
/translation="MNKSRQKELTRMLKQOSVISQRMNLISRLGFPVSGIILIAQWF
MARLIQIMIMENIPREALLPETLVTFVLAWVWLREBYAGAHIRPAIRFQV
LDRLOQGPAMIOGPPAGSNATLVLEBODHDVYARLPOMALAVSPLIIVAYFP
SNMAAILILGTAFLILPFAALVCMGADANRRFPLAARSGFPLDRMERTIRP
GRGAETESTRSASEDFRQRTWEVRLAFLSSGILEFTSLISLAVYFGSYGEL
DFGHYDGVTLAAGFLALILAPFFQPLRDGTGYHAKAQAAGADSILKTMETPLAH
PQGEAEALADPLTIEAEDLITSPSGKTAGLANTPLPAGQPAVLVGRSGKSL
LNAISGFSLYSGSLRINGIELRDSPSMRKHLVWGQNPQLPAKLRDNLARPDA
SEGLQALADNAWVSFPLIPQGVDPVPGQARLSGQORVAVVARALNPSLIL
LDPEASLIDHASEQVWMEALNALSRLQTLMTWQLEDLADMDVIMWQDGRITIEGR
YALSVAGFPATLAAHQEST"
/complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/complement(5083..6048)
/gene="trxB"
/locus_tag="S0888"
/function="enzyme; 2-Deoxyribonucleotide metabolism"
/note="residues 1 to 321 of 321 are 95.95 pct identical to residues 1 to 321 of 321 from Escherichia coli K-12 : B0888"
/codon_start=1
/transl_table=11
/product="thioredoxin reductase"
/protein_id="Aap16352.1"
/db_xref="GI:30040621"
/translation="MGTTKSKLILGSGPAYTAAYAAVANLOPVLTGMEKGGL
TTTTEVENNEDPDNDLGPILMARMEHATFEFEIIPDHINKVDLQNRPRLANGNG
EYTDALITATGASARYLGPSEBAFGKGVSACTDGFYRNQKVAIVGGATYAV
EALYLSINIASVHILHRRDGRRAKILIKRLMDVVENGNILHTNRKLEVTGQMGV
TGVLRDTONSDNIESLDVAGLFVAIGHSPTALFEQLELENGYIKVQSGIHGNATQ
TSIPGVPAAGVMDHIYRQAITSAGTCMALDAERYLIDGLADAK"
/6592..7086
/gene="trp"
/locus_tag="S0889"
/6592..7086
/gene="trp"
/locus_tag="S0889"
/function="regulator; Central intermediary metabolism: Pool, multipurpose conversions"
/note="residues 1 to 164 of 164 are 100.00 pct identical to residues 1 to 164 of 164 from Escherichia coli K-12 : B0889"
/codon_start=1
/transl_table=11
/product="regulator for leucine (or trp) regulon and high-affinity branched-chain amino acid transport system"
/protein_id="Aap16353.1"
/db_xref="GI:30040622"
/translation="MVDSSKRPGLDRIIDRIINLEKQGRISNVELSKRVGSPFP
CLEVRRLAEOGFIQGTALINPILADSLALPVEITLNGAPVPFPOFNAVOKLE
IOECHLVSGPDIYLLKTRVDMXVIRKLGLSETLRLRGVNDRTYVMEVKQSNRLV
IKTR"
/7221..11249
/gene="ftek"
/locus_tag="S0890"
/7221..11249
/gene="ftek"
/locus_tag="S0890"
/function="phenotype; Cell division"
/note="residues 1 to 737 of 737 are 68.24 pct identical to residues 1 to 737 of 1329 from Escherichia coli K-12 : B0890"
```

```

/codon_start=1
/transl_table=1
/product="cell division protein"
/protein_id="AAP16354.1"
/db_xref="GI:30040623"
/translation="MSQETREKKEVTLTSSGRLLLEALLTLVLPVWMLALISF
NPSPSPSOTAMHEPIHNGMFMALADTLFFIGVAVYTPVITVGGCFAPMRHS
SDEVIDEYFASLRITIGVALILITSCGLAINADIDIFYASGVGSLSTLQFARHS
SGGTIALLCVWAGLTLFTGMSVITIAELGIMINILITFASNTRRRDNDVDEYE
DDEBEDENHNGKHESRRARILRGALARRKLAELKINPMGRDNDALFSGKRDDE
EITTYAGVAAADDDVLFSGNRATOEYDEYELINSAPIREPVAVAAATTAQVSA
AVEPVTQTPVAVSVDPPOPTVAMOPPGGCGSEPIIAPRPGYQOQVAPVAVO
YNEPLOCVPOOPPYAPAPAEOPAOOPYAPAPRBPVAGNMAOEEOQSTFAPSTQ
TEQTYOOPPAEELVQOPDUEOCVVEPVEPVEPVEPVEPVEPVEPVEPVEPVEP
OLAAWQPIPEPKPEPKIKSLKAPVAAVPAVAAVPAVAAVPAVAAVPAVAAVPAV
TVAPVAFSLANGSGRPQVKGIGPOLRPRIRIVPTRELASYSGLPSPORAEKRA
REARNOQVDSQVNDDELDAQODELARQFQTOOQORGEYOHDVDPVNAEDADAA
EALAKQFAQTQOQRYSGEOPAGANPESLDDFEPSPMKALLDQDHPHPLTPIVEVQ
QPOQVAPQOQYQOQOQVAPQOQYQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQ
APQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQVAPQOQV
DLTTPPESEVEPTFALROMARLVEARLADFRKADLVNYSQPVITREINLAQV
KAARISNLSRLARSLTVAIRVAVVPIGKRYVGLTELPNKROTIVLRREVLDNAKPRD

```

Query Match
Best Local Similarity 98.0%; Score 340.8; DB 1; Length 292504;
Matches 345; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

```

QY 105 CGGTAAATATAGCGGCCCAATTCGTGAGCTGAACATTACAGTACGCTGGCGGTAACTC 164
DB 212560 CAGTAAATATAGCGGCCCAATTCGTGAGCTGAACATTACAGTACGCTGGCGGTAACTC 1212619
QY 165 TGCACCTTGCTGCAAACTGATGCCCGTAACCTTGACTATATACCCAGCATGGCGG 224
DB 212620 TGCACCTTGCTGCAAACTGATGCCCGTAACCTTGACTATATACCCAGCATGGCGG 212679
QY 225 CGGTAAATGTCAGATGTTGTCAGGGCTCAGATGACGCTCAATTCGATCCCAACG 284
DB 212680 TGGTAATGTCAGATGTTGTCAGGGCTCAGATGACGCTCAATTCGATCCCAACG 212739
QY 285 TGGCTTGCGTAAACGCGCTTACTTGTATCAGTGAACCGCAAAATTCGTAAGAACGCGT 344
DB 212740 TGGCTTGCGTAAACGCGCTTACTTGTATCAGTGAACCGCAAAATTCGTAAGAACGCGT 212799
QY 345 TAAACAGTTGCGTGTGGGCAACGCGCTCAGTGAACCGCAAAATTCGTAAGAACGCGT 404
DB 212800 TAAACAGTTGCGTGTGGGCAACGCGCTCAGTGAACCGCAAAATTCGTAAGAACGCGT 212859
QY 405 CAACGTGACTCAGGTGGCTTGTGTAACACGCGACCGCTCATGACTAA 456
DB 212860 CAACGTGACTCAGGTGGCTTGTGTAACACGCGACCGCTCATGACTAA 212911

```

RESULT 12
AF237726/c
LOCUS AF237726 437 bp DNA linear BCT 25-MAY-2000
DEFINITION Shigella flexneri insertion sequence IS600, complete sequence.
ACCESSION AF237726
VERSION AF237726.1 GI:7582371
KEYWORDS
SOURCE Shigella flexneri
ORGANISM Shigella flexneri
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Shigella.
1 (bases 1 to 437)
Sakellaris, H., Hamnik, N. K., Rajakumar, K., Bulach, D., Hunt, M.,
Sasakawa, C. and Adler, B.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10816548
2 (bases 1 to 437)
Sakellaris, H., Hamnik, N. K., Rajakumar, K., Bulach, D., Hunt, M.,
Sasakawa, C. and Adler, B.

TITLE Direct Submission
JOURNAL Submitted (23-FEB-2000) Microbiology, Monash University, Wellington
FEATURES
source
1. 437
/organism="Shigella flexneri"
/mol_type="genomic DNA"
/strain="SBA1100"
/serotype="2a"
/db_xref="taxon:623"
320. 392
/organism="Shigella flexneri"
/mol_type="genomic DNA"
/strain="SBA1100"
/serotype="2a"
/db_xref="taxon:623"
/insertion_seq="IS600"
/note="similar to IS600 deposited in GenBank Accession
Number X05952"
join(1. 319,393. 437)
/note="csg locus; similar to Escherichia coli K12 csg
locus deposited in GenBank Accession Number X90754"

Query Match
Best Local Similarity 98.1%; Score 311.4; DB 1; Length 437;
Matches 315; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 105 CGGTAAATATAGCGGCCCAATTCGTGAGCTGAACATTACAGTACGCTGGCGGTAACTC 164
DB 321 CAGTAAATATAGCGGCCCAATTCGTGAGCTGAACATTACAGTACGCTGGCGGTAACTC 262
QY 165 TGCACCTTGCTGCAAACTGATGCCCGTAACCTTGACTATATACCCAGCATGGCGG 224
DB 261 TGCACCTTGCTGCAAACTGATGCCCGTAACCTTGACTATATACCCAGCATGGCGG 202
QY 225 CGGTAAATGTCAGATGTTGTCAGGGCTCAGATGACGCTCAATTCGATCCCAACG 284
DB 201 TGGTAATGTCAGATGTTGTCAGGGCTCAGATGACGCTCAATTCGATCCCAACG 142
QY 285 TGGCTTGCGTAAACGCGCTTACTTGTATCAGTGAACCGCAAAATTCGTAAGAACGCGT 344
DB 141 TGGCTTGCGTAAACGCGCTTACTTGTATCAGTGAACCGCAAAATTCGTAAGAACGCGT 82
QY 345 TAAACAGTTGCGTGTGGGCAACGCGCTCAGTGAACCGCAAAATTCGTAAGAACGCGT 404
DB 81 TAAACAGTTGCGTGTGGGCAACGCGCTCAGTGAACCGCAAAATTCGTAAGAACGCGT 22
QY 405 CAACGTGACTCAGGTGGCTTGTGTAACACGCGACCGCTCATGACTAA 425
DB 21 CAACGTGACTCAGGTGGCTTGTGTAACACGCGACCGCTCATGACTAA 21

```

RESULT 13
STAJ2301
LOCUS STAJ2301 5103 bp DNA linear BCT 15-NOV-2000
DEFINITION Salmonella typhimurium csgC, csgF, csgB, csgD, csgB, csgA, and csgE genes.
ACCESSION AJ002301
VERSION AJ002301.1 GI:2739232
KEYWORDS csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
Romling, U., Bian, Z., Hammar, M., Sieralta, W. D. and Normark, S.
Curli fibers are highly conserved between Salmonella typhimurium
and Escherichia coli with respect to operon structure and
regulation
J. Bacteriol. 180 (3), 722-731 (1998)
JOURNAL MEDLINE 98117058

PUBMED 9457880
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling U.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
SOURCE Location/Qualifiers
1. 5103
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
1. 322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translation="CTLSGMVMAVLIHSGALAPHLEIVSVLTGIIVAFIMCIQAKQ
LILSPFPTFGACATFAGQDMRLVLPALGLIFGYAMKNSGIMLASRREQHSANT
AVTK"
complement(361..1194)
/gene="cs9g"
complement(361..1194)
/gene="cs9g"
/function="11poprotein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="cs9g protein"
/protein_id="CA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translation="MPRLILVAILLSGCTAPPKQAAPKPTLMPRAOSYKDLTHLP
PTGKIPFVSVINIODETGOKFPPASNPSTAPASATMALTKDSRFITLEQYGLQ
NLNLRKIIIRAQENGTAVAMNRIPLQSLTANIMVEGSIIGYSNVAASGGVARG
IGADTQYQLDIAVLRLVNVSTGEIISVNTSKTIIISYEVOAGVFRIIDYQRLLEG
IGYTSNPEVMCLMSALETGVIPLINDIDRGLMDLQWKADRONDIIVKYNELSVPE
S"
complement(1221..1637)
/gene="cs9f"
complement(1221..1637)
/gene="cs9f"
/codon_start=1
/transl_table=1
/product="cs9f protein"
/protein_id="CA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translation="MRVHAVALMLFSPLTWAGNMTFOFRNPNFGGNPNNSFLNS
AQASGYKDPAYNDFGIETPSALDNFOATIOSQILGLLNTNITGKRGAVNTDFII
DIARBDQLOLVNTDRKTRGRTIEVSGLOEQSTDP"
complement(1664..2059)
/gene="cs9g"
complement(1664..2059)
/gene="cs9g"
/codon_start=1
/transl_table=1
/product="cs9g protein"
/protein_id="CA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translation="MKRYLTWIVAELLFATGNLHANEVEVEPGILTDHTVSSIGHE
FYRPSDKSESEYGNLTINERPARSGSWTTITVNDVIFQTFLEPKMDFEKTIVF
ALAQTERALNRQIDOTILSTSDLARDEF"
complement(2064..2714)
/gene="cs9d"
complement(2064..2714)
/gene="cs9d"

/function="transcriptional regulator"
/codon_start=1
/transl_table=1
/product="Cs9d protein"
/protein_id="CA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translation="WFNEVSSHGHTLLITKPSLQATALLQHLKOSLATITGKLNIIQ
RSLIEDISAGCIVLMDMEADKLIHYQDNLSRNNNIKTLLNTPDPYPRRIENMP
HINGVFATQEDQEHVSGLOGILRGECYFSQKLASVLIITHSGNRYNSTEASALLTHRE
KEILNKLRTIGSNNEIARSLFISENTVKTLYNLFKKIAYKGRFOAVSWANDNLR"
complement(2869..2896)
/gene="cs9dBERg"
complement(2869..2874)
/gene="cs9dBERg"
complement(2891..2896)
/gene="cs9dBERg"
3344..3376
/gene="cs9BA"
3344..3349
/gene="cs9BA"
3371..3376
/gene="cs9BA"
3469..3924
/gene="cs9BA"
3469..3924
/gene="cs9B"
3469..3924
/gene="cs9B"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cs9B protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKRKLPMHLITIGAGIATATYTDLARSFYRANVELSSRN
QAATIGQVGTDSNARVQESKLSVTISOEGNNRRAVDQAGNTNFIYIEGTGNADA
SISQSAVGNAAAIQKSGKNKANTOYGTAVVVOKSHMARVQR"
3469..3531
/gene="cs9B"
3532..3921
/gene="cs9B"
/product="Cs9B protein"
/function="nucleator"
3966..4421
/gene="cs9A"
3966..4421
/gene="cs9A"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="cs9A protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/db_xref="SWISS-PROT:P55225"
/translation="MKLKVAFALIVSGSALAGVVPKMGGGGNHGGNSGPDST
LSIVQGSANAALALOSDARKESTTISGCGNCAVDGCAQNSSTIELTONGPNNAT
IDQMANNSDITVQIGGNALVNOTASDSSVMVROVGFNNATANQY"
3966..4025
/gene="cs9A"
4026..4418
/gene="cs9A"
/product="Cs9A protein"
/function="fiber subunit"
4483..4809
/gene="cs9C"
4483..4809
/gene="cs9C"
/codon_start=1
sig_peptide
mat_peptide
sig_peptide
mat_peptide
gene
CDS
sig_peptide
mat_peptide
gene
CDS
sig_peptide
mat_peptide
gene
CDS

```

/transl_table=1
/product="CsgC Protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translation="MHTLLIAASNOIFFTTQGGDIYTVIPOVTLNPPCCVOVIL
SVDDVGQGGSHQKOTLPLPANOPLETSLRSLSVNTSSDSVKIIVTVSDGSLHLSQO
WPSAQ"
4483..4533
/gene="csgC"
/mat_peptide
4534..4806
/gene="csgC"
/product="CsgC"
/gene="csgC"
/gene="ORF103"
/gene="ORF103"
/codon_start=1
/transl_table=1
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translation="MSVKKKNIPIAIGLCIAFFTHSAVGGQTVGGVHFRGAIVEPL
CDISTHANENI"

```

Query Match 56.8%; Score 259.2; DB 1; Length 5103;
 Best Local Similarity 73.0%; Pred. No. 2.9e-64; Mismatches 123; Indels 0; Gaps 0;
 Matches 333; Conservative 0;

```

OY 1 ATGAACTTTTAAAGTAGAAGCAATTCGCGCAATGTAATTCCTCCGGTAGCCCTCGCA 60
DB 3966 ATGAACTTTTAAAGTAGAAGCAATTCGCGCAATGTAATTCCTCCGGTAGCCCTCGCA 60
OY 61 GGTGTGTCTCTCAAGTGGCGGGCGGGTAAACACGATGTTGGGGTAAATATAGCCGC 120
DB 4026 GGGCTGTCTCAAGTGGCGGGCGGGTAAACACGATGTTGGGGTAAATATAGCCGC 120
OY 121 CCAATTTGAGCTGGAACATTTACGATGATGCGGTGGGTAATCTGCACTTGTCTGCA 180
DB 4086 CCGGATTTCCAGCTTGAACATTTATCGATGATGCGGTGGGTAATCTGCACTTGTCTGCA 180
OY 181 ACTGATGCGCGTAACTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4145
DB 4146 AGGATGCGCGTAACTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4145
OY 241 GTTGTCTGAGCTGGAACATTTACGATGATGCGGTGGGTAATCTGCACTTGTCTGCA 4205
DB 4206 GTAGGCGGAGGTCGGATTAACGATGATGATGATGATGATGATGATGATGATGATGAT 4205
OY 301 GGTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4265
DB 4266 GGTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4265
OY 361 GGCACGCTGCTGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4325
DB 4326 AATAACGCGGCGGCTGATTAACGATGATGATGATGATGATGATGATGATGATGATGAT 4325
OY 421 GCGTTTGTAAACGCGGCGCTCATCTACTACTAA 456
DB 4386 GGTTTTGGCAACGCGGCGCTCATCTACTACTAA 4421

```

RESULT 14
 LOCUS AE008749
 DEFINITION *Salmonella typhimurium* LT2, section 53 of 220 of the complete
 genome.
 ACCESSION AE008749 AE006468
 VERSION AE008749.1 GI:16419641
 KEYWORDS
 SOURCE *Salmonella typhimurium* LT2
 ORGANISM *Salmonella typhimurium* LT2

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS
 CONSRTM
 TITLE
 JOURNAL
 COMMENT

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; *Salmonella*.
 1 (bases 1 to 22411)
 McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Latreille, P., Courtney, L., Portolillo, S., All, J., Dante, M., Du, F.,
 Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Stoeckert, C., Niu, N., Mahoney, E., Ryan, E., Sun, H., Florea, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LT2
 Nature 413 (6858), 852-856 (2001)
 2 (bases 1 to 22411)
 15677609

coding sequences below are predicted from manually evaluated
 computer analysis, using similarity information and the programs;
 GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
 GeneMark; <http://opal.biology.gatech.edu/GeneMark/>
 R numbers were kindly provided by Junko Yabuzaki and the Kyoto
 Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
 and Pedro Romero and Peter Karp at EcolCyc;
<http://ecocyc.org/PangeaSystems.com/ecocyc/>
 The analyses of ribosome binding sites and promoter binding sites
 were kindly provided by Heladia Salgado, Julio Collado-Vides and
 http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameweb

FEATURES
 source
 gene
 RBS
 CDS

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by sequence
 from more than one m13 subclone.
 Location/Qualifiers
 1..22411
 /organism="Salmonella typhimurium LT2"
 /mol_type="genomic DNA"
 /strain="LT2; SGGC 1412; ATCC 700720"
 /db_xref="ATCC:700720"
 /db_xref="taxon:99287"
 /note="LT2"
 434..1308
 /gene="phoH"
 /note="synonym: STM1126"
 434..439
 /gene="phoH"
 /note="putative RBS for phoH; RegulonDB:STWS1H001398"
 454..1308
 /gene="phoH"
 /note="similar to E. coli phoB-dependent, ATP-binding pho
 regulon component; may be helicase; induced by p
 starvation (AAC74105.1); bla_{SE} hit to AAC74105.1 (354
 aa), 92% identity in aa 71 - 354"
 /codon_start=1
 /transl_table=11
 /product="phoB-dependent ATP-binding pho regulon
 component"
 /protein_id="AAU20057.1"
 /db_xref="GI:16419642"
 /translation="MCAHNRVQENAGDYLQKYSMGKQAVYKRRREAKVLRDS
 RSHKRESESVTSLVQGVGVEATGARDSDSPYKARERQAHYLAIDSQILFAT
 GEAGCGKWTISAKAABALIHQVERIYTRPVLDDEDLGFLPDIDIAEKFAVFRPV
 YDVLKRLGASFMQYCLREIGKVEIAPAYRGRFENAVVILDEAQNVTAAQMKMF

Db 18069 GCCACCATGACGACGAGTGAACGCTAAACCTCCGATATTACTGTCGTCATACATACGCGCGT 18128
Qy 361 GGACAGCGTCTGCTGACTGTTGACGAGACTGATCTAATCTCTCCGTCAGACTAGCTAGGTT 420
Db 18129 AATPACCGCGCGCTGCTGTTATACAGCCGATCATGATTCAGGCTAATGTCGCTCAGGTT 18188
Qy 421 GGCTTTGTGTACACGACGACCGCTGCTACTACTACTAA 456
Db 18189 GGTTTTGGACAAACCCACGCTAACCAATATTA 18224

RESULT 15
AL627269

LOCUS 254050 bp DNA linear BCT 04-JUL-2003
DEFINITION Salmonella enterica serovar typhi (Salmonella typhi) strain CT18,
complete chromosome; segment 5/20.
ACCESSION AL627269 AL513382
VERSION AL627269.1 GI:16502231
KEYWORDS
SOURCE
ORGANISM Salmonella enterica subsp. enterica serovar typhi
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS Parkhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,
Sebania, M., Baker, S., Basham, K. L., Bentley, S. D., Holden, M. T. G.,
Compton, P., Cronin, A., Davis, P., Davies, R. M., Dowd, L., White, N.,
Jagels, K., Krog, A., Larsen, T. S., Leather, S., Mout, S., O'Garra, P.,
Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J.,
Stevens, K., Whitehead, S. and Barrett, B. G.
Complete genome sequence of a multiple drug resistant Salmonella
enterica serovar typhi CT18
Nature 413 (6858), 848-852 (2001)
21534947
11677608
2 (bases 1 to 254050)
Parkhill, J.
Direct Submission
Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
Hinxton, Cambridgeshire CB10 1SA, UK
E-mail: parkhill@sanger.ac.uk
COMMENT
Notes:
Details of S. typhi sequencing at the Sanger Centre are available
on the World Wide Web.
location: /www.sanger.ac.uk/Projects/S_typhi/.
FEATURES
source
1. 254050
/organism="Salmonella enterica subsp. enterica serovar
typhi"
/mol_type="genomic DNA"
/strain="CT18"
/db_xref="taxon:90370"
181. 2793
/gene="STY1078"
/note="synonym: pepN"
181. 2793
/gene="STY1078"
/note="Orthologue of E. coli pepN (AMPN_ECOLI); Faeta hit
to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap"
/codon_start=1
/transl_table=11
/product="aminopeptidase N"
/protein_id="CAD08183.1"
/db_xref="GI:16502232"
/db_xref="GOA:O82770"
/db_xref="SPTREMBL:O82770"
/translation="MTQDQAKRHDYADYQITDIEITDLDIAEKTVAISQAVR
HAPADPILDESDLTIVSIHVNDAPWTAKEEGALITSLPERFTLRIVNEISPA
NTRLEIYSGDLCTQCEAEGRHITWYLDPRDVLARFTTKIADSKYPELISN
RVAQGLENGRWVQMDPFPKCYVLFALVAGDVLADFTTTSRGEVLELYVDG

misc_feature
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
family M1, score 245.10, E-value 1.9e-84"
1060. 1089
/gene="STY1078"
/note="PS00142 Neutral zinc metalloproteases,
zinc-binding region signature"
3001. 4011
/gene="STY1079"
/note="synonym: PYRD"
3001. 4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:O82789"
/db_xref="SWISS-PROT:O82789"
/translation="MYDPRVAKALFQDPRAEHEFTQRLRTGTPLAEVRYQVPT
KPYTCMGITFKNPGLAAGLDKDEICIDALGANGFSLGELGTVPBPQNDKRLPR
LVDAAGLNRGPNFNNLVNDLVENVKKAHFDGLIGINIGKNDLPVNGKDDVLTKE
KVAVAGYIAINISSPNTPGRLTLOYDADLDLLTAIKKNDLOVTHKKYVVAIKI
APDICEELIQVADSLRNHIDGVIATNTTLDLSIVQMGNCOCOTGSLGRPLQKST
EIRRLSLELKGQPLIIGVGIDSVIAAREKINAGNTVQIYISGFIKGPLIKELVT
HI"
3136. 3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODehase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238. 3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874. 3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174. 4719
/gene="STY1080"
4174. 4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:O82788"
/translation="WMARKPDNNRWYVDEEDHDMWLLDNGMLFRSFRSRKMLTPDA
GEMTPTGIGDASVWLSDTAEQVNLIVVEGENAALCTLAAPGVLAGRTMGDAIK
IMNDLKPQVCHSFRSLEQAV"
complement (4716. 5825)
/gene="STY1081"
complement (4716. 5825)
/gene="STY1081"

misc_feature
/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase
family M1, score 245.10, E-value 1.9e-84"
1060. 1089
/gene="STY1078"
/note="PS00142 Neutral zinc metalloproteases,
zinc-binding region signature"
3001. 4011
/gene="STY1079"
/note="synonym: PYRD"
3001. 4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=11
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:O82789"
/db_xref="SWISS-PROT:O82789"
/translation="MYDPRVAKALFQDPRAEHEFTQRLRTGTPLAEVRYQVPT
KPYTCMGITFKNPGLAAGLDKDEICIDALGANGFSLGELGTVPBPQNDKRLPR
LVDAAGLNRGPNFNNLVNDLVENVKKAHFDGLIGINIGKNDLPVNGKDDVLTKE
KVAVAGYIAINISSPNTPGRLTLOYDADLDLLTAIKKNDLOVTHKKYVVAIKI
APDICEELIQVADSLRNHIDGVIATNTTLDLSIVQMGNCOCOTGSLGRPLQKST
EIRRLSLELKGQPLIIGVGIDSVIAAREKINAGNTVQIYISGFIKGPLIKELVT
HI"
3136. 3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODehase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238. 3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874. 3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174. 4719
/gene="STY1080"
4174. 4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:O82788"
/translation="WMARKPDNNRWYVDEEDHDMWLLDNGMLFRSFRSRKMLTPDA
GEMTPTGIGDASVWLSDTAEQVNLIVVEGENAALCTLAAPGVLAGRTMGDAIK
IMNDLKPQVCHSFRSLEQAV"
complement (4716. 5825)
/gene="STY1081"
complement (4716. 5825)
/gene="STY1081"

gene
CDS

```
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/translation="MATHSLRFLHPVKSMDIGLTHALADISGLAFDRIFMWTSDGT
PITARQPMWRFTPSPLHDLGLHLPAGSSALVAFDPTQDAPTVEWGNHFTFRAVA
PTAINOMLSGFSPSDVQLCWGQPOLTRAVKHNAPLGFADGYPYLINBSLRDLOQ
RCPAGVQMEORFNPVLVSGVAAMEDSKVRIGDVIPTDVVKPCSRGCIPTTSPKQO
KHPSGEPLATQAPRTIADNGDVPGQVLIARNSGVIVGCEVELIATAPAKAGATT
LDSTVTPKHPDGSATTTIDQGTFCGNNQVYLBQLENGKIRIPSCAGICGCCRRIR
LLEBSVSPKSKASAGDDITLSCSPKFTALRLN"
/complement(4737..4955)
/gene="STY1081"
/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
/gene="STY1082"
5924..8032
/note="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY_HAEIN (P44524; P43945) Faeta
scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/translation="
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82786"
/db_xref="SPTREMBL:Q82786"
/translation="MNSLFASTAGLELLKIELEKAGVCGQVVGVRHQCPTRLI
YOSIMSRILASRIILPMGECKVYSDLDLYFVQAINWTEIFNPGATFVHFSGLNDTI
RNSQYGAKVDAIVDAFTRNLPRPNVDRSPDLRIWMLNKETASTALDSGDGLH
LRGYRDRGLAPIKETLAAALVMRSQMPGTPPLDPMCGSGTLLIEAMWATDRAPGL
HRGHMFGSGMAHDEALWQEVAAQOTRAKGLAESHSFPGSDDAVIERASNNAR
RAGIGELITFEVKQVAQSNPLPKPGYGTVISNPGERLSDPALIALHSLIGRTMK
NORGWNLISFASPDLLGSLQADQOFKAKNGPLDVCQNYHIAETTASKPATYA
EDYANRLRKNIKKLEKAROREICTRLYDADLPENYAVDRIGMAYTOEYAPKTV
DAQARORLFDLIAATLSVLGIPIKVLKTRERQKQNOYQKMSKEGFELESEYNA
RLWNLIDYLDPTGLFDHRIARMLGEMSKGDFLNFYSYGSASVHAGLGARNTT
VDMSRTYLEMAERNILNGLSGRAHRLIOADCLMAREANQFDLIFIDPTFNSKR
MEESPVDORHDVALMKDLKRLRKGRTIMFNNKGRPMDEGLAELGLTNGEITQKT
LSPPARNROIHNCMLIRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
5506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJJK_ECOLI (554 aa), 34% identity in 524 aa
overlap
```

```
Db 88906 ATGAACCTTTTAAAGTAGAAGCAATTCGACCAATCGTAGTTTCTGCGAGGCTCTGACT 88965
QY 61 GGTGTGTTCTCTAGTACGGCGGCGCGGTAAACACGGTGTGGCGGTAAATAGCGGC 120
Db 88966 GCGCGTTCACATATGGGGGGGGGGGTATATATACGGCGGGCAATAGTTCCGGG 89025
QY 121 CCAATTCTAGCTGAACATTATACAGTACGGTGGCGGTAACTCTGCACTTCTCTGCA 180
Db 89026 CCGATTTCACAGTTGAGCATTTATACAGTACGGTTCGGTAAACGTCGCCCTTCTCTGCA 89085
QY 181 ACTATGCCGTAACTCTGACTTGAATTAACCCAGATGGCGGCGGTATGGTGCAGAT 240
Db 89086 AGCGATGCCCTTAATCTGAACACACATTAACAGACGGTTATGTGAACGGCGGAT 89145
QY 241 GTTGATCAGGCTCAGATGACGATCAATCGATCTGACCCACGTGCGCTTCGTAAACAGC 300
Db 89146 GTAGGCCAGGGTGGGATTAACAGTACTATTAAGTGAATGCTGAAATGGTTTCAGAAACAT 89205
QY 301 GCTACTCTTATCATGTGGAACGGCAAAATTTCTGAATGACGGTTAAACAGTTGCGTGT 360
Db 89206 GCCACCAATCGACAGGTGAACGTAAACCTCCGATATTACTGTCGTCGTCATTAACGGCGGT 89265
QY 361 GGCAACGGTGTGCGAGTTGACCAAGATGATCTGAACCAAGTGGCTTCGTAAACAGT 420
Db 89266 AATAACCGCGCGCTGTTAATCAACCGCATCTTAATTCACCGTAAATGGTTCGTAGSTT 89325
QY 421 GCGTTTGTAAACACGACCGCTCATCATGACTTAA 456
Db 89326 GTTTGGCAACACGCGCAGGCTTAACAGTATTAA 89361
```

Search completed: March 15, 2004, 22:49:48
Job time : 1966.17 secs

```
Query Match 56.5%; Score 257.6; DB 1; Length 254050;  
Best Local Similarity 72.8%; Pred. No. 1.5e-63;  
Matches 332; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
QY 1 ATGAACCTTTTAAAGTAGAAGCAATTCGACCAATCGTAGTTTCTGCGAGGCTCTGCA 60
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds

(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456
Sequence: 1 atgaacttttaaaagtaga.....cgaccgcctcactgactactaa 456Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq.*
6: /cgn2_6/ptodata/2/ina/bacfilest.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	254.4	55.8	456	1	US-08-233-788A-58 Sequence 58, Appl
2	173.4	38.0	361	1	US-08-233-788A-56 Sequence 56, Appl
3	35.6	7.8	1788	4	US-09-252-991A-4632 Sequence 4632, Ap
4	35.6	7.8	2250	4	US-09-252-991A-4181 Sequence 4181, Ap
5	35.6	7.8	3102	4	US-09-252-991A-4429 Sequence 4429, Ap
6	34	7.5	552	4	US-09-252-991A-2024 Sequence 2024, Ap
7	34	7.5	669	4	US-09-252-991A-1852 Sequence 1852, Ap
8	34	7.5	1185	4	US-09-252-991A-1931 Sequence 1931, Ap
9	33.4	7.3	4403765	3	US-09-103-840A-2 Sequence 2, Appl
10	33.4	7.3	4411529	3	US-09-103-840A-1 Sequence 1, Appl
11	33	7.2	1830121	4	US-08-557-884-1 Sequence 1, Appl
12	33	7.2	1830121	4	US-08-643-990A-1 Sequence 1, Appl
13	32	7.0	2993	2	US-08-225-488-6 Sequence 6, Appl
14	31.8	7.0	4403765	3	US-09-103-840A-2 Sequence 2, Appl
15	31.4	6.9	1317	4	US-09-252-991A-10989 Sequence 29, Appl
16	31.2	6.8	1810	4	US-09-667-135-29 Sequence 586, App
17	31	6.8	1422	4	US-09-328-352-586 Sequence 75, Appl
18	30.6	6.7	435	3	US-08-569-147-75 Sequence 190, App
19	30.6	6.7	1923	4	US-09-056-556-190 Sequence 185, App
20	30.6	6.7	1923	4	US-09-072-596-185 Sequence 190, App
21	30.6	6.7	1923	4	US-09-072-596-190 Sequence 190, App
22	30.6	6.7	4411529	3	US-09-103-840A-1 Sequence 1, Appl
23	30	6.6	6122	3	US-08-403-545-1 Sequence 1, Appl
24	30	6.6	6122	3	US-08-404-381-1 Sequence 1, Appl
25	30	6.6	1230025	4	US-09-158-452A-1 Sequence 915, App
26	29.8	6.5	771	4	US-09-489-039A-915 Sequence 393, App
27	29.8	6.5	1128	4	US-09-328-352-393

28	29.8	6.5	2652	4	US-09-489-039A-5584 Sequence 57, Appl
29	29.8	6.5	87563	4	US-09-453-702B-57 Sequence 106, App
30	29.2	6.4	14654	4	US-08-961-577-106 Sequence 8302, App
31	28.8	6.3	429	4	US-09-252-991A-8302 Sequence 7, Appl
32	28.8	6.3	912	2	US-08-993-228-7 Sequence 8312, App
33	28.8	6.3	1023	4	US-09-328-352-3184 Sequence 8312, App
34	28.8	6.3	1395	4	US-09-252-991A-8312 Sequence 6856, App
35	28.8	6.3	9507	4	US-09-489-039A-6896 Sequence 5247, App
36	28.8	6.3	12720	1	US-08-403-866-11 Sequence 3008, App
37	28.6	6.3	217	4	US-09-313-294A-5247 Sequence 9, Appl
38	28.6	6.3	693	4	US-09-328-352-3008 Sequence 2442, App
39	28.6	6.3	1507	4	US-09-453-323-1 Sequence 32, Appl
40	28.6	6.3	1873	3	US-09-461-474-9 Sequence 4713, App
41	28.4	6.2	1380	4	US-09-489-039A-2442 Sequence 317, App
42	28.4	6.2	62909	4	US-09-596-002-32 Sequence 317, App
43	28.2	6.2	298	4	US-09-313-294A-4713 Sequence 317, App
44	28.2	6.2	346	3	US-09-060-756-317 Sequence 317, App
45	28.2	6.2	346	4	US-09-670-314-317

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Sharon S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
TITLE OF INVENTION: OF SALMONELLA
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 55.8%; Score 254.4; DB 1; Length 456;
Best Local Similarity 72.4%; Pred. No. 3.3e-70;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTAGAAGCAATTCGAGCAATCTATTCTCCGATAGCGCTCTGGCA 60
DB 1 ATGAACCTTTAAAGTAGAAGCAATTCGAGCAATCTATTCTCCGATAGCGCTCTGGCA 60
QY 61 GGTGTGTTCTCTCAGTAGACGGCGCGCGGTAAACGAGTGTGTGGGTAAATTAACGGC 120
DB 61 GGTGTGTTCTCTCAGTAGACGGCGCGCGGTAAACGAGTGTGTGGGTAAATTAACGGC 120
QY 121 CCAATTTCTGAGCTGAACATTTTACAGTACGCTGCGGTAACTCTGACTGTCTGCAA 180
DB 121 CCGAGCTCAAGCTTATGAGCATTTATCAGTACGCTGCGGTAACTCTGACTGTCTGCAA 180
QY 181 ACTGATGCGCGGTAACTCTGAGTAACTCCAGCAGTGTGCGCGGTAAATGTGCAAT 240
DB 181 ACGGATGCGCGGTAACTCTGAGTAACTCCAGCAGTGTGCGCGGTAAATGTGCAAT 240
QY 241 GTTGTGAGGCTCAAGTGAAGTCAATGATCTGACCCAGTGTGCGGTAAACGCGCT 300
DB 241 GTAGGCGAGGCTGAGTAAATGATCTGATTAATGATCTGACCCAGTGTGCGGTAAACGCGCT 300
QY 301 GCTACTCTTATGATGAGTGAACGCGCAAAATTCGAAATGACGCTTAAACAGTTGCGTGT 360
DB 301 GCGACCATTCACACAGTGAACGCTTAAATTCGAAATGACGCTTAAACAGTTGCGTGT 360
QY 361 GCGAACGCTGCTGAGTGAACGAGCTGATCTGCTCCGTCAGCGTCAAGTCAAGCT 420
DB 361 AATTAAGCGCGCTGCTGATTAATGAGCAGCATCTGATCTCAAGCTGATGCTGCGT 420
QY 421 GCGTTGTGTAACAGCGACCGCTCATGACTACTAA 456
DB 421 GCGTTGTGTAACAGCGACCGCTCATGACTACTAA 456

```

RESULT 2

```

US-08-233-788A-56
/ Sequence 56, Application US/08233788A
/ Patent No. 5635617
/ GENERAL INFORMATION:
/ APPLICANT: Doran, James L.
/ APPLICANT: Kay, William W.
/ APPLICANT: Collinson, Karen S.
/ APPLICANT: Clouthier, Sharon C.
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
/ NUMBER OF SEQUENCES: 61
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Seed and Berry
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: U.S.A.
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patent In Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/233,788A
/ FILING DATE: 26-APR-1994
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: King, Joshua
/ REGISTRATION NUMBER: 35,570
/ REFERENCE/DOCKET NUMBER: 920043.403C2
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ TELEX: 372836 SEEDANBERY
/ INFORMATION FOR SEQ ID NO: 56:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 361 base pairs

```

```

/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..357
/ US-08-233-788A-56

```

```

Query Match
Best Local Similarity 38.0%; Score 173.4; DB 1; Length 361;
Matches 234; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

```

```

QY 64 GTTGTCTCTGATGAGCGCGCGGTAAACGAGTGTGTGGGTAAATTAACGGCCCA 123
DB 1 GTGTAACCAAGTGGGCGCGCGGTAAATTAACGAGTGTGTGGGTAAATTAACGGCCCA 123
QY 124 AATTGTGAGCTGAACATTTTACAGTACGCTGCGGTAACTCTGCACTTGTGCAACT 183
DB 61 GACTCAACGTTGAGCATTTTATGAGTACGCTGCGGTAACTCTGCACTTGTGCAACT 183
QY 184 GATGCGCGTAACTCTGACTTGAATTAACGAGTGTGCGCGGTAAATGTGCAATGTT 243
DB 121 GATGCGCGTAACTCTGACTTGAATTAACGAGTGTGCGCGGTAAATGTGCAATGTT 243
QY 244 GGTGAGGCTGAGTGAACGCTCAATGATCTGACCCAGTGTGCGGTAAACGCGCT 303
DB 181 GCGCAGGCTGAGTGAATTAATGATTAATGATCTGACCCAGTGTGCGGTAAACGCGCT 303
QY 304 ACTCTTGTATCAGTGAACGCGCAAAATTCGAAATGACGCTTAAACAGTTGCGTGTGC 363
DB 241 ACCATGACAGTGAACGCTTAAATTCGAAATGACGCTTAAACAGTTGCGTGTGC 363
QY 364 AACGCTGCTGAGTGAACGAGCTGATCTGACTACTAA 456
DB 301 AACGCTGCTGAGTGAACGAGCTGATCTGACTACTAA 456

```

RESULT 3

```

US-09-252-991A-4632/C
/ Sequence 4632, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ PRIOR FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 4632
/ LENGTH: 1788
/ TYPE: DNA
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-252-991A-4632

```

```

Query Match
Best Local Similarity 7.8%; Score 35.6; DB 4; Length 1788;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

```

```

QY 30 AGCAATCGTATTCTCCGATGAGCGCTGCGAGTGTGTTCTCTGATGCGCGCGCGCG 89
DB 1391 AGCGGTGTATTGCGCGGTGCGGATGCGCGGATGCGAGAAAGCGCGCGCG 1332
QY 90 TAACACGCTGCTGCGGTAT 111
DB 1331 TGATCAGGTGCGGTAGCAT 1310

```

RESULT 4
US-09-252-991A-4181
; Sequence 4181, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4181
; LENGTH: 2250
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4181

Query Match 7.8%; Score 35.6; DB 4; Length 2250;
Best Local Similarity 64.6%; Pred. No. 0.26;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 30 AGCATGCTATTCTCCGCTAGCGCTCTGCGAGCTTTGTTCTCTCAGTACGGCGCGCGCGG 89
Db 339 AGCGGCTATTCTCCGCTAGCGCTCTGCGAGCTTTGTTCTCTCAGTACGGCGCGCGCGG 398
Qy 90 TAACCAAGCTGTGCGCGGTAAT 111
Db 399 TGATCAGGCTGCGCGTGAACAT 420

RESULT 5
US-09-252-991A-4429/c
; Sequence 4429, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 4429
; LENGTH: 3102
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4429

Query Match 7.8%; Score 35.6; DB 4; Length 3102;
Best Local Similarity 64.6%; Pred. No. 0.3;
Matches 53; Conservative 0; Mismatches 29; Indels 0; Gaps 0;
Qy 30 AGCATGCTATTCTCCGCTAGCGCTCTGCGAGCTTTGTTCTCTCAGTACGGCGCGCGCGG 89
Db 1411 AGCGGCTATTCTCCGCTAGCGCTCTGCGAGCTTTGTTCTCTCAGTACGGCGCGCGCGG 1352
Qy 90 TAACCAAGCTGTGCGCGGTAAT 111
Db 1351 TGATCAGGCTGCGCGTGAACAT 1330

RESULT 6
US-09-252-991A-2024/c
; Sequence 2024, Application US/09252991A

Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 2024
; LENGTH: 552
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-2024

Query Match 7.5%; Score 34; DB 4; Length 552;
Best Local Similarity 56.1%; Pred. No. 0.44;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 147 GTACGCTGCGCGCTAGCTCTGCTCTGCTCAACTGATGCCGCTAGCTGACTTAC 206
Db 119 GTGCGCGCGCGCTAGCTCTGCTCTGCTCAACTGATGCCGCTAGCTGACTTAC 60
Qy 207 TATTACCAAGCTGTGCGCGGTAATGTGTCAGATGTTGTCAGGCTTCAGATGA 260
Db 59 TGTGCTCTTGTGCGCGCTGTTGTCAGATGTCGCGGCTGTCAGGCTTCAGATGA 6

RESULT 7
US-09-252-991A-1852
; Sequence 1852, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 1852
; LENGTH: 669
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-1852

Query Match 7.5%; Score 34; DB 4; Length 669;
Best Local Similarity 56.1%; Pred. No. 0.48;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 147 GTACGCTGCGCGCTAGCTCTGCTCTGCTCAACTGATGCCGCTAGCTGACTTAC 206
Db 462 GTGCGCGCGCGCTAGCTCTGCTCTGCTCAACTGATGCCGCTAGCTGACTTAC 521
Qy 207 TATTACCAAGCTGTGCGCGGTAATGTGTCAGATGTTGTCAGGCTTCAGATGA 260
Db 522 TGTGCTCTTGTGCGCGCTGTTGTCAGATGTCGCGGCTGTCAGGCTTCAGATGA 575

RESULT 8
US-09-252-991A-1931
; Sequence 1931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.

```

1  TITLE OF INVENTION:  DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
2  TITLE OF INVENTION:  TUBERCULOSIS
3  FILE REFERENCE:  24366-20007 .00
4  CURRENT APPLICATION NUMBER:  US/09/103,840A
5  CURRENT FILING DATE:  1998-06-24
6  NUMBER OF SEQ. ID NOS:  2
7  SOFTWARE:  PatentIn Ver. 2.1
8  SEQ ID NO 1
9
10 LENGTH: 4411529
11
12 TYPE:  DNA
13
14 ORGANISM:  Mycobacterium tuberculosis
15
16 OTHER INFORMATION:  H37Rv
17
18 US-09-103-840A-1

```

Query Match	7.3%	Score 33.4	DB 3	Length 441529
Best Local Similarity	72.9%	Pred. No. 30		
Matches 43	Conservative 0	Mismatches 16	Indels 0	Gaps 0

QY 206 CATTACCGCATGGGGGCTATGTGCAGATGTCACAGGCTCAGATGACAGC 264
|||||
Dd 371189 CTGTGGCGACGACGACCGCGATCTGGTGGCGAGTGGCGAGGCTGAATGAGAGC 371247
|||||

RESULT 11
US-03-557-884-1
Sequence 1, Application US/09557884
Patent No. 6,955,784

APPLICANT: Fleischmann et al.
 TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof
 NUMBER OF SEQUENCES: 1
 CORRESPONDENCE ADDRESSES:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: MD
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3 1/2 inch diskette
 COMPUTER: Dell Pentium
 OPERATING SYSTEM: MS DOS v6.22
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/557, 884
 FILING DATE: 25-Apr-2000
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/476,102
 FILING DATE: JUN-5-1995

NAME: MICHELLE S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH:

```

? ? TYPE: nucleic acid
? ? STRANDEDNESS: double
? ? TOPOLOGY: linear
? ? SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Query Match          7 2%   Score 33; DB 4; Length 1830121;
Best Local Similarity 55.8%; Pred. No. 30;
Matches    63; Conservative    0; Mismatches    50; Indels    0; Gaps    0

```

122 CAAATCTGAGCTGAACATTTACAGTAGCGGTGGCGTAATCTGCACCTTGCTCTGCAA 181

Db 740925 CGATTTTTCGCAAGATCGTTTAAATGATATGCGCTTACAAATTCCTCGTTGTTATGACACA 740984

Qy 182 CTGATGCCCGTAACTCTGACTTGTATTTACCCAGCATGGCGCGCTAATGCT 234

Db 740985 TTGATGCTGTGTTTAAACCAAAATTATTTGAAATGTCGGCGTTAAGGCT 741037

RESULT 12

US-09-643-990A-1

Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams

Owen White
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville,

STATE: MD

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette

COMPUTER: Dell Pentium

OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-AUG-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:

NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186PICI

TELECOMMUNICATION INFORMATION:

TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-643-990A-1

Query Match 7.2%; Score 33; DB 4; Length 1830121;

Best Local Similarity 55.8%; Pred. No. 30;

Matches 63; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 122 CAAATTCGAGCTGAACATTTTACACGATCGGTGGCGGTAATGCTCTGCTGCAA 181

Db 740925 CGATTTTTCGCAAGATCGTTTAAATGATATGCGCTTACAAATTCCTCGTTGTTATGACACA 740984

Qy 182 CTGATGCCCGTAACTCTGACTTGTATTTACCCAGCATGGCGCGCTAATGCT 234

Db 740985 TTGATGCTGTGTTTAAACCAAAATTATTTGAAATGTCGGCGTTAAGGCT 741037

RESULT 13

US-08-225-488-6/c
Sequence 6, Application US/08225488

Patent No. 5846802

GENERAL INFORMATION:

APPLICANT: Buxton, Frank

APPLICANT: Hinnen, Albert

APPLICANT: Visser, Jacob

TITLE OF INVENTION: No. 5846802el Fungal Protease

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Ciba-Geigy Corporation

STREET: 7 Skyline Drive

CITY: Hawthorne

STATE: NY

COUNTRY: USA

ZIP: 10532

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/225,488

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/047,214

FILING DATE: 13-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Spruill, W. Murray

REGISTRATION NUMBER: 32,943

REFERENCE/DOCKET NUMBER: 4-19055/R/CIP

TELECOMMUNICATION INFORMATION:

TELEPHONE: 919-541-8615

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2993 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Aspergillus niger

STRAIN: M400

IMMEDIATE SOURCE:

CLONE: pTZPBD

FEATURE:

NAME/KEY: promoter

LOCATION: 1..829

FEATURE:

NAME/KEY: CDS

LOCATION: join(830..1153, 1205..1649, 1697..1785, 1841

LOCATION: 2233)

OTHER INFORMATION: /note="subtilisin-type protease

OTHER INFORMATION: PEPP of Aspergillus niger; product of gene pepd"

FEATURE:

NAME/KEY: intron

LOCATION: 1154..1204

FEATURE:

NAME/KEY: intron

LOCATION: 1650..1696

FEATURE:

NAME/KEY: intron

LOCATION: 1786..1840

US-08-225-488-6

Query Match 7.0%; Score 32; DB 2; Length 2993;

Best Local Similarity 46.4%; Pred. No. 3.9;

Matches 104; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

QY 65 TTGTTCTCAGTACGGGGGGGCGGTAAACAGGTGTGGCGGTAAATTAAGCGCCCA 124
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2181 TTGTAGCCGGAAGATTGGGGCTACCGCCACATTGTGATACAGCATTCGCGTAGCGCAAC 2122
QY 125 ATTCTGAGCTGAACATTATACAGTAAGGTGGCGGTAACTCTGCACTGTCTGCAAACTG 184
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2121 CTCTTAGCTCGGTGCTGTGAGCGCGTGGGGTAGCAAGTCCCGCAAGCCCATCAAAATG 2062
QY 185 ATGCCCGTAACTCTGACTTATTAACCAAGCAGCGCGGTAAATGGTGCAGATGTTG 244
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2061 AGGATTAACCTGTCAATGAGGTGAGCCATGACGTGCCGAGATCGTTGTGGTGGCC 2002
QY 245 GTCAGGGCTCAGATGACAGTCAATGATCTGACCCCAAGTGGC 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2001 GAGTTGAGCGCGTCCATGACAGAAAGTACTTGTCTCTCCCGGGGC 1958

```

```

RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
US-09-103-840A-2

```

```

Query Match 7.0%; Score 31.8; DB 3; Length 4403765;
Best Local Similarity 46.0%; Pred. No. 72;
Matches 108; Conservative 0; Mismatches 127; Indels 0; Gaps 0;
QY 82 GCGCGCGGTAAACAGGTGTGGCGGTAAATTAAGCGCCCAATTCTGAGCTGAACATT 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4029096 GCGGTGGGGGTGACGCTGTGAGCGGGGTAAACGCGGGATCGCGGTGACGCGCGCC 4029037
QY 142 TACCAAGTACGGTGGCGGTAACTCTGCACTTGTGCAAACTGATGCCGTAATCTGAC 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4029036 GCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCTTTCGCGCTCTCGGTAGCTCCGGG 4028977
QY 202 TTGACTATTAACCAAGTGGCGCGGTAAATGGTGCAGATGTTGTCAGAGGCTCAGATGAC 261
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4028976 CTGGGCGGTGCGCGGTGAGCGGGGTAAATGGCGGGCGGTGAGCCGCGTGTGGCC 4028917
QY 262 AGCTCAATCATGATGACCAACGATGCTTGGGTAAACAGCGCTACTCTTATCATGCT 316
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4028916 GCGTCCCGCGCGGTGGGCGCTGCGGCGCGGTGCGCGCAACCTCGCGCAGT 4028862

```

```

RESULT 15
US-09-252-991A-10989/c
; Sequence 10989, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

```

```

; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10989
; LENGTH: 1317
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10989

```

```

Query Match 6.9%; Score 31.4; DB 4; Length 1317;
Best Local Similarity 52.7%; Pred. No. 4.2;
Matches 68; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 64 GTTGTCTCAGTACCGCGCGGTAAACAGGTGTGGCGGTAAATTAAGCGCCCA 123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 513 GTTGTCTCAGTACCGCGCGGTAAACAGGTGTGGCGGTAAATTAAGCGCCCA 454
QY 124 AATTCTGAGCTGAACATTATACAGTAAGGTGGCGGTAACTCTGCACTGTCTGCAAACT 183
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 GCGACTGCGGAGCGCTTGTCAAGTCTCGGCGGTGGGAGGTGTGTACTGACGCG 394
QY 184 GATGCCCGT 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 393 GATGCCCGT 385

```

Search completed: March 16, 2004, 04:36:54
Job time: 72.6647 secs


```

Qy 107 GTAATATATAGGGCCCAATTCTGAGCTGAACATTACAGTACCGTGGCGTAACCTCTG 166
Db 70182 GTGGTATGATGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 166
Qy 167 CACTTGCTCTGCAAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70123
Db 70122 ATGGTCGACACCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 226
Qy 227 GTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 70066
Db 70065 ATTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 286
Qy 287 GCTTCGCTGACAGCGCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 70006
Db 70005 GTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 346
Qy 347 AACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69946
Db 69945 GTGATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 69909

```

RESULT 2

```

US-10-017-161-1781
; Sequence 1781, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRABANT, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1781
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-017-161-1781

```

Query Match

Best Local Similarity 8.6%; Score 39; DB 14; Length 1390; Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```

Qy 62 GTGTTGTTCTCAGTACGCGCGCGGTAACACCGTGTGCGGTAAATATATACGGCC 121
Db 853 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
Qy 122 CAAATTCGAGCTGAACATTATACAGTACGCGGTAACTCTGCACTTCTGCAAA 181
Db 913 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
Qy 182 CTGATGCGCGTAACCTGACTTGAATTAACCAAGATGCGCGGTAAATGATGATGATGATG 972
Db 973 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 241
Qy 242 TTGTCAGGCGCTCAGATGACAGCTCAATGATCTGACCAAGTGGCTTGGTAACAGCG 1032
Db 1033 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Qy 302 CTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Db 1093 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1152

```

```

Qy 362 GCAACGCTGCTGACGTTGA 380
Db 1153 GCAATGTGTGTATGATGATG 1171

```

RESULT 3

```

US-10-292-798-1437
; Sequence 1437, Application US/10292798
; Publication No. US2003023583A1
; GENERAL INFORMATION:
; APPLICANT: SUMA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABRABANT, HIROYUKI
; TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
; FILE REFERENCE: 084335/166
; CURRENT APPLICATION NUMBER: US/10/292,798
; PRIOR FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: 10/017,161
; PRIOR FILING DATE: 2001-12-18
; PRIOR APPLICATION NUMBER: JP 2001-246789
; NUMBER OF SEQ ID NOS: 2070
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1437
; LENGTH: 1390
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; LOCATION: source
; FEATURE:
; LOCATION: (1)..(1390)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (201)..(1190)
US-10-292-798-1437

```

Query Match

Best Local Similarity 8.6%; Score 39; DB 15; Length 1390; Matches 144; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

```

Qy 62 GTGTTGTTCTCAGTACGCGCGCGGTAACACCGTGTGCGGTAAATATATACGGCC 121
Db 853 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 912
Qy 122 CAAATTCGAGCTGAACATTATACAGTACGCGGTAACTCTGCACTTCTGCAAA 181
Db 913 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 972
Qy 182 CTGATGCGCGTAACCTGACTTGAATTAACCAAGATGCGCGGTAAATGATGATGATGATG 972
Db 973 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 241
Qy 242 TTGTCAGGCGCTCAGATGACAGCTCAATGATCTGACCAAGTGGCTTGGTAACAGCG 1032
Db 1033 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
Qy 302 CTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1092
Db 1093 GTGATGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
Qy 362 GCAACGCTGCTGACGTTGA 380
Db 1153 GCAATGTGTGTATGATGATG 1171

```

RESULT 4

```

US-10-029-386-24675
; Sequence 24675, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.

```

```
APPLICANT: Rank, David R.
TITLE OF INVENTION: HANZEL, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
FILE REFERENCE: AEMICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 24675
LENGTH: 625
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AP001754.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.8
OTHER INFORMATION: EST HUMAN HIT: B1015471.1, EVALUATE 1.00e-03
US-10-029-386-24675

Query Match      8.4%; Score 38.2; DB 14; Length 625;
Best Local Similarity 46.2%; Pred. No. 0.026;
Matches 162; Conservative 0; Mismatches 188; Indels 1; Gaps 1;

QY 29 CAGCAATGATATTCCTCCGTAAGCGCTCTGCGACAGTGTGTTCTTCAGTAACGGCGGCGGCG 88
DB 87 CATCACTCTTACTTATGAGTAGTGGGCGCGGTGGTGGATGATGATGATGATGATGATGATGATG 146
QY 89 GTACACGAGGTGAGTGGCGGTATATATAGCGCCCAATTCAGAGTGAACATTTACCGAGT 148
DB 147 ATGGTATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 205
QY 149 ACGGTGCGGTACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 208
DB 206 GATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 265
QY 209 TTACCCAGACATGCGCGCGGTAAATGTCAGATGTTGTCAGAGGCTCAGATGACAGCTCA 268
DB 266 ATGGTATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 325
QY 269 TCGATCTGACCCCAAGTGGCTTCCGTACACGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 328
DB 326 GTGATGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 385
QY 329 ATTCTGAATGACGCTTAAACAGTTCGGTGGTGGCAACGCTGCTGCTGCTGCTGCTGCTGCTG 379
DB 386 ATGTTGATGCTGATGCTTGGCAGTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 436

RESULT 5
US-10-017-161-1981/c
Sequence 1981, Application US/10017161
Publication No. US20030143668A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1981
LENGTH: 1972
TYPE: DNA
ORGANISM: Homo sapiens
```

```
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1972)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(1772)
FEATURE:
NAME/KEY: modified base
LOCATION: (975)..(994)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified base
LOCATION: (1319)..(1328)
OTHER INFORMATION: a, t, c, g, unknown or other
US-10-017-161-1981

Query Match      8.2%; Score 37.4; DB 14; Length 1972;
Best Local Similarity 49.2%; Pred. No. 0.085;
Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 62 GTGTTGTTCTCTCAGTACGCGCGCGGTAAACACGCGTGGCGGTAAATATAGCGGCC 121
DB 460 GTGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 401
QY 122 CAATTCGAGCTGAACATTTACAGTACGCTGGCGGTAACTGCACTTGTCTGCAAA 181
DB 400 ATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 341
QY 182 CTGATGCGCGGTACTCTGACTTGAATTAACCCAGACATGCGCGGTAAATGTCAGATG 241
DB 340 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 281
QY 242 TTGCTCAGGCTCAGATGA 260
DB 280 GTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 262

RESULT 6
US-10-292-798-1629/c
Sequence 1629, Application US/10292798
Publication No. US2003023583A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: AKIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: GUANOSINE TRIPHOSPHATE-BINDING PROTEIN COUPLED RECEPTORS
FILE REFERENCE: 084335/166
CURRENT APPLICATION NUMBER: US/10/292,798
CURRENT FILING DATE: 2002-11-13
PRIOR APPLICATION NUMBER: 10/017,161
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: JP 2001-246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2070
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1629
LENGTH: 1972
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
LOCATION: source
FEATURE:
LOCATION: (1)..(1972)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(1772)
FEATURE:
NAME/KEY: modified base
LOCATION: (975)..(994)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
```

LOCATION: (1319)..(1328)
 OTHER INFORMATION: a, c, g, unknown or other
 US-10-292-798-1629

Query Match
 Best Local Similarity 8.2%; Score 37.4; DB 15; Length 1972;
 Matches 98; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 62 GTGTGTCCTCCAGTACGCGCGCGGTACCAACCGGTGTGGCGGTATATATACGGCC 121
 DB 460 GTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 401
 QY 122 CAATTTTGAAGCTGAAATTAACAGTACGCGGTGTGGCGGTATATATATACGGCC 181
 DB 400 ATGCTGT 341
 QY 182 CTGATGCCGTAACTCTGACTTACTATTAACCAAGCGCGGTGTGGCGGTATATATAC 241
 DB 340 GTGATGT 281
 QY 242 TTGGTCAGGCTTCAGATGA 260
 DB 280 GT 262

RESULT 7

US-09-738-626-1857
 Sequence 1857, Application US/09738626
 Publication No. US20020197605A1
 GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIALI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 PRIOR FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 PRIOR FILING DATE: 2000-08-03
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 1857
 LENGTH: 2580
 TYPE: DNA

ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1857

Query Match
 Best Local Similarity 8.1%; Score 37; DB 9; Length 2580;
 Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 77 ACGCGCGCGCGGTAAACCAAGTGTGTGGCGGTAAATATAGCGCCCAATTTCTGAGCTGA 136
 DB 1046 AGGGTGAAGCGTGGCGCGATGGGAGAGCGCGCGCGATGATGATGATGATGATGATG 1105
 QY 137 ACATTTTACAGTACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1105
 DB 1106 ACACGAGCAAGT 193
 QY 194 ACTCTGACTTACTATTAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1165
 DB 1166 CATTATTAAGGCGT 1225

QY 254 CAGATGACAGCTCAATGATCTGACCCAGAGTGGCTTCGGTAAACGCGCTACTTGTATC 313
 DB 1226 CAGATGCGAGTCCATTAAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1285
 QY 314 AGTGAAGCGCAAAATTTCTGAATGACGGTTAAACAGTTTGGTGTGGCAAC 366
 DB 1286 TGTGTTTCTGAAAGCGCATGAAATCAATCAACAAGGGGTGAGAGGGCGAC 1338

RESULT 8

US-09-738-626-1/c
 Sequence 1, Application US/09738626
 Publication No. US20020197605A1
 GENERAL INFORMATION:

APPLICANT: NAKAGAWA, SATOSHI
 APPLICANT: MIZOGUCHI, HIROSHI
 APPLICANT: ANDO, SEIKO
 APPLICANT: HAYASHI, MIKIRO
 APPLICANT: OCHIALI, KEIKO
 APPLICANT: YOKOI, HARUHIKO
 APPLICANT: TATEISHI, NAKO
 APPLICANT: SENOH, AKIHIRO
 APPLICANT: IKEDA, MASATO
 APPLICANT: OZAKI, AKIO

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 FILE REFERENCE: 249-125
 CURRENT APPLICATION NUMBER: US/09/738,626
 PRIOR FILING DATE: 2000-12-18
 PRIOR APPLICATION NUMBER: JP 99/377484
 PRIOR FILING DATE: 1999-12-16
 PRIOR APPLICATION NUMBER: JP 00/159162
 PRIOR FILING DATE: 2000-04-07
 PRIOR APPLICATION NUMBER: JP 00/280988
 NUMBER OF SEQ ID NOS: 7059
 SOFTWARE: PatentIn ver. 3.0
 SEQ ID NO 1
 LENGTH: 3309400
 TYPE: DNA

ORGANISM: Corynebacterium glutamicum
 US-09-738-626-1

Query Match
 Best Local Similarity 8.1%; Score 37; DB 9; Length 3309400;
 Matches 140; Conservative 0; Mismatches 150; Indels 3; Gaps 1;

QY 77 ACGCGCGCGCGGTAAACCAAGTGTGTGGCGGTAAATATAGCGCCCAATTTCTGAGCTGA 136
 DB 1784428 AGGGTGAAGCGTGGCGCGATGGGAGAGCGCGCGATGATGATGATGATGATGATG 1784369
 QY 137 ACATTTTACAGTACGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 193
 DB 1784368 ACACGAGCAAGT 1784309
 QY 194 ACTCTGACTTACTATTAACCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1784309
 DB 1784308 CATTATTAAGGCGT 1784249
 QY 254 CAGATGACAGCTCAATGATCTGACCCAGAGTGGCTTCGGTAAACGCGCTACTTGTATC 313
 DB 1784248 CAGATGCGAGTCCATTAAGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1784189
 QY 314 AGTGAAGCGCAAAATTTCTGAATGACGGTTAAACAGTTTGGTGTGGCAAC 366
 DB 1784188 TGTGTTTCTGAAAGCGCATGAAATCAATCAACAAGGGGTGAGAGGGCGAC 1784136

RESULT 9

US-10-029-386-25133/c
 Sequence 25133, Application US/10029386
 Publication No. US20030194704A1
 GENERAL INFORMATION:

```

; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: clone ID: UC-2MFLB7312F11 FLI

```

```

;
; APPLICANT: LOSTOH, Camt
;
; TITLE OF INVENTION: QUORUM SENSING SIGNALING IN BACTERIA
;
; FILE REFERENCE: UIZ-038CP
;
; CURRENT APPLICATION NUMBER: US/10/389,647

```

```

RESULT 15
US-10-029-386-7202
/ Sequence 7202, Application US/10029386
/ Publication NO. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEOTIC ACID PROBES USEFUL FOR G
FILE REFERENCE: A60MICA-X-2
CURRENT APPLICATION NUMBER: US/10/029,386
CURRENT FILING DATE: 2001-12-20
NUMBER OF SEQ ID NOS: 34288
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 7202
LENGTH: 532
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC003693.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3

```


THIS PAGE BLANK (USPTO)

FEATURES

http://xenopus.nibb.ac.jp.
Location/Qualifiers

1.680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:9355"
/clone="X186B22"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match

Best Local Similarity 99.4%; Score 453.4; DB 12; Length 680;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGAACCTTTAAAGTAAAGCAATTCAGACATCGTATTCGCGTAGCCCTCTGGCA 60
593 ATGAACCTTTAAAGTAAAGCAATTCAGACATCGTATTCGCGTAGCCCTCTGGCA 60
61 GGTGTGTTCTCTCAAGTACGGCGCGGCGGTAAACCAAGCGGTGCGGTAAATAGCGGC 474
533 GGTGTGTTCTCTCAAGTACGGCGCGGCGGTAAACCAAGCGGTGCGGTAAATAGCGGC 474
121 CCAATTCCTGAGTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGTCTGCA 180
473 CCAATTCCTGAGTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGTCTGCA 180
181 ACTGATCCCGTAACTCTGACTGACTTGAATTAACCGACATGGCGCGGTAAATAGTGCAGAT 240
413 ACTGATCCCGTAACTCTGACTGACTTGAATTAACCGACATGGCGCGGTAAATAGTGCAGAT 240
241 GTTGTGACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
353 GTTGTGACGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
301 GCTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294
293 GCTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
361 GGCACACGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
233 GGCACACGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
421 GGCCTTGTGAACACGCGACCGCTCATCGTACTGATGATGATGATGATGATGATGATGAT 174
173 GGCCTTGTGAACACGCGACCGCTCATCGTACTGATGATGATGATGATGATGATGATGAT 138

RESULT 2
CD374421/c 558 bp mRNA linear EST 19-AUG-2003
LOCUS TNMwmc2H373 Wuchereria bancrofti microfilaria cDNA (SAM9551L-WbMc)
DEFINITION Wuchereria bancrofti cDNA clone TNMwmc2H3 5', mRNA sequence.
ACCESSION CD374421
VERSION CD374421.1 GI:31229939
KEYWORDS EST.
SOURCE Wuchereria bancrofti
ORGANISM Wuchereria bancrofti
Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;
Onchocercidae; Wuchereria.
1 (bases 1 to 558)
Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.
Toward the characterization of the Wuchereria bancrofti
microfilarial transcriptome with comparisons to those of Brugia
malayi and Onchocerca volvulus
Unpublished (2003)
Contact: Thomas B. Nutman
Laboratory of Parasitic Diseases
NIH
Building 4 Room 126, Bethesda, MD 20892-0425, USA
Email: tnutman@niaid.nih.gov

JOURNAL
COMMENT

FEATURES

Seq primer: pJ1nuescript T3.
Location/Qualifiers

1.558
/organism="Wuchereria bancrofti"
/mol_type="mRNA"
/db_xref="taxon:6293"
/clone="TNMwmc2H3"
/dev_stage="microfilaria"
/lab_host="X11-Blue MRP"
/clone_lib="Wuchereria bancrofti microfilaria cDNA (SAM9551L-WbMc)"
/note="Vector: Lambda Uni-ZAP XR; Site 1: Eco RI; Site 2: Xho I; Lymphatic filarial nematode parasite of humans, isolated from the blood of an infected individual from Guyana, South America and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 2.9 x 10⁵ independent recombinants and the average insert size is ~1kb. The library was constructed by Sandra U. Laney. The library is available from Dr.S.A.Williams, email: genome@smith.edu."

ORIGIN

Query Match

Best Local Similarity 13.3%; Score 60.8; DB 14; Length 558;
Matches 93; Conservative 0; Mismatches 25; Indels 2; Gaps 2;

63 TGTGTGCTCTGACAGTACGGCGCGGTAAACCAAGCGGTGCGGTAAATAGCGGCC 121
174 TGTGTGCTCTGACAGTACGGCGCGGTAAACCAAGCGGTGCGGTAAATAGCGGCC 115
122 CAAATTCGAGTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGTCTGCA 180
114 CAAATTCGAGTGAACATTTACAGTACGGTGGCGGTAACTGTGCACTTGTCTGCA 55

RESULT 3
B3348812/c 607 bp mRNA linear EST 06-MAR-2002
LOCUS B3348812 Dictyostelium discoideum cDNA library, Af Dictyostelium
DEFINITION B3348812 Dictyostelium discoideum cDNA clone dda34011 3', mRNA sequence.
ACCESSION B3348812
VERSION B3348812.1 GI:19219319
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
1 (bases 1 to 607)
Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
Full length cDNA of Dictyostelium discoideum at the aggregation
stage
Unpublished (2002)
Contact: Tadashi Shin-i
Center for Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.

FEATURES

Location/Qualifiers

1.607
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda34011"
/sex="mat A"
/dev_stage="aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, Af"

ORIGIN

Query Match

9.5%; Score 43.4; DB 12; Length 607;

```

/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/strin="Columbia 0"
/db_xref="taxon:3702"
/clone="GK-160B03-013216"
/clone_lib="Arabidopsis thaliana T-DNA insertion lines"
/notes="PCR was performed on DNA from Arabidopsis thaliana plants (T1) which were transformed with the T-DNA from vector PAC161. The lines contain one or more T-DNA insertions. The DNA fragment(s) resulting from the PCR were directly sequenced to determine the genomic sequence flanking the insertion. Sequences displaying significant similarity to the A. thaliana nuclear genome sequence were processed for submission. T-DNA derived sequences were

```

Email: schabell@iastate.edu

Individual basecall and confidence value were assigned using the Phred software (<http://www.phrap.org/>). Overall sequence quality assessment and vector trimming was conducted using the Lucy software (version 1.16a, <http://www.cigr.org/softlab/>). Lucy parameters were set to ensure an overall trimmed quality of 97.5% or better without any vector fragments in the chosen high quality region of each sequence. Low-quality bases between the poly-T and the high-quality region were replaced with N's to serve as spacers using a Perl program (`test_process.pl`), written by Dr. Hui-Hsien Chou.

PCR Primers

FORWARD: primer T7-1 (AA TAC GAC TCA CTA TAG)
BACKWARD: primer Sp6 (ATT TAA GTG ACA CTA TG)
Seq primer: universal (GTA AAG CGA CGG CCA GT)
POLYA=Yes.

Location/Qualifiers

1. 540

/organism="Zea mays"

/mol_type="mRNA"

/culivar="B73"

/db_xref="taxon:4577"

/clone="MEST54-D07"

/rbase_type="mixed"

/lab_host="DH10B"

/clone.lib="ISUM6"

/note=Vector: pSlip7 (4.43 kb); Site 1: EcoRI; Site 2: NotI; Tissue samples were collected and partially pooled prior to RNA extraction. First-strand cDNAs were prepared from 21 individual pools of oligo-dT selected mRNAs by priming with 21 different NotI oligo-dT tag primers (5'-NACTGGAAGATTGGCGCCGCGCCNNNNNTTTTCTTTTCTTTT-3'). Distinguishingable 'bar code' tags, (N)6, were used for each separate first-strand cDNA synthesis. Hence, these bar code tags can be used to identify the mRNA pool from which a particular cDNA clone was derived. The 'bar code' tags associated with specific tissue sources are:

ATAAGC--Germinated seeds and seedlings (1, 2, 8, 11 DAG);

ACTGCG--Mixed mature tissues (17, 21, 38, 69, 77 DAG);

CACGAC--Kernels (3, 5, 10, 15, 20, 25, 30, DAG);

TACCC--Adventitious roots (65 DAG); CAGCG--Tassels (3-39 cm, 53 and 56 DAG); AGGTAC--Husks (73 DAG); GACAC--Silks; 56, 59 DAG); TGAGCG--Immature ears (0.2-3.0 cm, 53, 56, 59 DAG);

ATAAGC--unpollinated first ears; CTAAGC--ear shanks; GTGAGC--Cycloheximide-treated callus; GTACAC--Anaerobic acid)-treated seedlings; GATGCC--Kinectin-treated seedlings; AAGACC--ACPC (1-aminocyclopropane-1-carboxylic acid)-treated seedlings; GCTCA--Brassinolide-treated seedlings; CTAGCC--ABA (Abscissic acid)-treated seedlings; TACCGA--GA (Gibberellic acid)-treated seedlings; GCAGGA--JA (Jasmonic acid)-treated seedlings; Equal amounts of first-strand cDNA from each reaction were combined and used as template for DNA PolI-catalyzed 2nd strand synthesis. After the addition of EcoRI adaptor, the ds-cDNAs were digested with NotI. Molecules between 0.5 and 2.0 kb were directionally cloned into the EcoRI and NotI sites of the pSlip7 expression vector. Plasmid DNA isolated from the library was digested with NotI to remove empty vector clones. Linear DNAs from 5.4 to 7 kb were gel purified and ligated at low concentration to promote recircularization. Ligation products were precipitated and transformed into DH10B host cells. The complexity of the resulting library was 1.2 x 10⁶."

ORIGIN

Query Match 8.2%; Score 37.2; DB 12; Length 540;
Best Local Similarity 55.4%; Pred. No. 23;
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 45 CGGTAGCGCTTGGCAGGTGTTTCTCAGTACGCGCGCGGTAAACCGGTG 104
DB 410 CGGTCAGGTCTCGCGCGCTTGGTGTCTCGCGCGCGCGCTTGGCGGTG 351
QY 105 CGGTATATATAGCGCGCCCAATTCTGAGCTGAACATTACAGTACGCGGTG 164
DB 350 CGGCTTGTGCGCGCTTGAAGATCGGCTTGGCGCGCGCGCTTGGCGGTG 291
QY 165 TGCACTTCT 174
DB 290 TGGGCTCGGT 281

RESULT 6 B2997124/c
LOCUS B2997124 771 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TB ZM.0.6-1.0 KB Zea mays genomic clone ZMBR1381104,
genomic survey sequence.
ACCESSION B2997124
VERSION B2997124.1 GI:29240541
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 771)

AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uteback, T.,
Rennick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUGGV50TB
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
source Location/Qualifiers
1..771
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMBR1381104"
/clone.lib="ZM.0.6-1.0 KB"
/note=Vector: PCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN

Query Match 8.2%; Score 37.2; DB 28; Length 771;
Best Local Similarity 50.6%; Pred. No. 27;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 193 AACTTGACTGACTATTAACCGATGCGCGGTATGGGAGATGTTGACAGGC 252
DB 374 AACACTTGCTGCTAGTATTAACCGATGCTGCTGATATATTAAGCTTACCGTTAA 315
QY 253 TAGATGACAGCTGATGATCTGACCAAGCTGCTTGGTAAACAGCTTCTGAT 312
DB 314 TATCAACGTTCCACTGATATTAACCAAGCTTCAATCGGTATTAACCAAGCTTCA 255
QY 313 CAGTGAACGGCAAAATTTGAATGACGCTTAACAGTTGCGGTGCGCAACGGT 370
DB 254 CGGTATATATGAAAGCTTGTGAGTAAATTAACCAAGCTTCCGCGGTATTAACGAG 197

RESULT 7 B2997127
LOCUS B2997127 797 bp DNA linear GSS 25-MAR-2003
DEFINITION PUGGV50TD ZM.0.6-1.0 KB Zea mays genomic clone ZMBR1381104,
genomic survey sequence.
ACCESSION B2997127
VERSION B2997127.1 GI:29240544
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoidae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 797)

AUTHORS

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Uteback, T.,
Rennick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Bennetzen, J.
TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PUGGV50TB
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
source Location/Qualifiers
1..797

/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA381104"
/clone_lib="ZM 0.6-1.0 KB"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

Query Match 8.2%; Score 37.2; DB 28; Length 797;
Best Local Similarity 50.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 193 AACTTGACTGACTTATTAACAGATGCGCGGTAATGTCAGATGTTGTCAGGCG 252
DB 505 AACCTTTCGTCAGTATTAACCAAGCGTTCGTAATTAATTCATGCTTACCGTTAA 564
QY 253 TCAGATGACAGCTCATGATCGATCGACCCAGCGTTCGTAACAGCGCTACTCTTAT 312
DB 555 TAATCCAAAGCTTCCACTAGTATTAACCAAGCGTTCGTAATTAACCCAGCTTCA 624
QY 313 CAGTGAACCGCAAAATTTCTGAATGACGTTAAACAGTTGCGTGGCAACGATG 370
DB 625 CGGTATATATGCAACGTTTCGTAGATTAATTAACCAACGCTTCCCGGTATTAACGAG 682

RESULT 8
LOCUS CG200536 818 bp DNA linear GSS 21-AUG-2003
DEFINITION PULS94TD ZM 0.6-1.0 KB Zea mays genomic clone ZMMBTA059019,
genomic survey sequence.

ACCESSION CG200536
VERSION CG200536.1 GI:34091597
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 818)

REFERENCE
AUTHORS Whitehead, C.A., Quackenbush, J., Van Aken, S., Utecher, T.,
Benwick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
Benner, J.

TITLE Maize Genomes Consortium
JOURNAL Unpublished (2003)
COMMENT Contact: Cathy Whitehead
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitehead@tigr.org
Seq primer: TF
Class: sheared ends.

FEATURES

SOURCE

1. 818
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA059019"
/clone_lib="ZM 0.6-1.0 KB"
/note="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high
cot selected genomic DNA library"

ORIGIN

Query Match 8.2%; Score 37.2; DB 29; Length 818;
Best Local Similarity 50.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 193 AACTTGACTGACTTATTAACAGATGCGCGGTAATGTCAGATGTTGTCAGGCG 252
DB 277 AACCTTTCGTCAGTATTAACCAAGCGTTCGTAATTAATTCATGCTTACCGTTAA 336

QY 253 TCAGATGACAGCTCATGATCGATCGACCCAGCGTTCGTAACAGCGCTACTCTTAT 312
DB 337 TAATCCAAAGCTTCCACTAGTATTAACCAAGCGTTCGTAATTAACCCAGCTTCA 396
QY 313 CAGTGAACCGCAAAATTTCTGAATGACGTTAAACAGTTGCGTGGCAACGATG 370
DB 397 CGGTATATATGCAACGTTTCGTAGATTAATTAACCAACGCTTCCCGGTATTAACGAG 454

RESULT 9
LOCUS B2115869/c 789 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-255D7.TVB CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-255D7, genomic survey sequence.

ACCESSION B2115869
VERSION B2115869.1 GI:23756816
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 789)
Zhao, S., Shetty, J., Shatsman, S., Tseng, G., Geer, K.,
Shvachbeyn, A., Gebregorgis, B., Overton, L., Russell, D., Chen, D.,
Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other GSSs: CH230-255D7.TU
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat330.htm). For BAC library
availability, please contact Pieter de Jong (pjejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Plate: 255 row: D column: 7
Seq primer: TV
Class: BAC ends.

FEATURES

SOURCE

1. 789
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SnHsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-255D7"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARBACT.3; Site 1: MboI; Site 2: MboI;
CHORI-230 Rat (BN/SnHsd/MCw) BAC library produced by
Pieter de Jong"

ORIGIN

Query Match 8.1%; Score 37; DB 28; Length 789;
Best Local Similarity 45.8%; Pred. No. 31;
Matches 127; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 46 GTAGCGCTTGGAGAGGTGTTCTCAGTACGCGGCGGTAACACAGCGTGTGTC 105
DB 680 GGTGTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 621
QY 106 GGTATATATAGCGGCCCAATTTCTGAGCTGAACATTAACAGTACGATGCGGTAACCT 165
DB 620 GTTGCTGATGTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 561

QY 166 GCATTGCTGCAACTGATGCGCCGTAAGTCTGACTTATTACCGAGATGCGGC 225
DB 560 GGTGATGCTGCTGCTGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 501
QY 226 GGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
DB 500 GGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 441
QY 286 GCTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 322
DB 440 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 404

RESULT 10
BX342087 1171 bp mRNA linear EST 01-MAY-2003
LOCUS BX342087 Homo sapiens HELA CELLS COT 25-NORMALIZED Homo sapiens
DEFINITION CDNA clone CS0DK009YH08 5-PRIME, mRNA sequence.
ACCESSION BX342087
VERSION BX342087.1 GI:30308886
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
COMMENT Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1171)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 2235.f. For more information about this cluster, see
http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DK009D040P1
cluster=2235.f. Contact : Feng Liang Email : fliang@lifetech.com
URL : http://fulllength.invitrogen.com/Invitrogen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DK009D040P1.

FEATURES
source
Location/Qualifiers
1..1171
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DK009YH08"
/cell_type="HELA CELLS COT 25-NORMALIZED"
/clone_id="HELA"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN
Query Match 7.9%; Score 36; DB 13; Length 1171;
Best Local Similarity 12.1%; Pred. No. 72;
Matches 22; Conservative 95; Mismatches 65; Indels 0; Gaps 0;

QY 71 CTACGTAACGGGCGGCGGTACCAACGCGGTGCGGTATTAATAGCGGCCAAATTTCTG 130
DB 947 CCCCCKSGGSGSKWAGMKCKCMGCKKMKKSKCMKMRKMGMBWGMGMMMKMKMY 1066
QY 131 AGCTGAACATTTTCCAGTACGCGGTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 190
DB 1007 KGGCGGAGMMSBGMATTTMMKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 1066
QY 191 GTAACCTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG 250
DB 1067 KKKKKGYNHMMKK 1126
QY 251 GC 252
::

DB 1127 SS 1128

RESULT 11
CB642466/c 675 bp mRNA linear EST 08-APR-2003
LOCUS CB642466/c
DEFINITION OSJNEB02K15.r OSJNEB Oryza sativa (japonica cultivar-group) CDNA
clone OSJNEB02K15 3', mRNA sequence.
ACCESSION CB642466
VERSION CB642466.1 GI:29637457
KEYWORDS EST.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
COMMENT Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 675)
AUTHORS Jantsen, Y., C., Lu, G., Gowda, M., Hatfield, J., Zhou, B., Mazur, E., Kudrna, D., Dean, R., Soderlund, C., Wing, R., and Wang, G.
TITLE Large-scale identification of ESTs involved in the interaction between rice and Magnaporthe grisea
JOURNAL Unpublished (2003)
COMMENT Contact: Rod Wing
Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210088, Tucson, AZ 85721-0088, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: http://genome.arizona.edu
PCR Primers
FORWARD: gta aac cga cgg cca gtc g
BACKWARD: gga aac agc tat gac cat g
Plate: 02 row: K column: 15
Seq primer: gga aac agc tat gac cat g.
Location/Qualifiers
1..675
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNEB02K15"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_id="OSJNEB"
/note="Vector: pBluescript II KS +; Site 1: EcoRI; Site 2: XhoI; 24 hrs after inoculation with Rice Blast (Che 86061)."

ORIGIN
Query Match 7.9%; Score 35.8; DB 14; Length 675;
Best Local Similarity 59.2%; Pred. No. 63;
Matches 61; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 6 ACTTTTAAAGTAGAGCAATTCAGCAATGCTATTCCTGCTGCTGCTGCTGCTGCTGCT 65
DB 626 ACTTCACACGGGTGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 567
QY 66 TCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 108
DB 566 GCTTCGAGACGGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 524

RESULT 12
CNS02APC/c 774 bp DNA linear GSS 01-SEP-2000
LOCUS CNS02APC/c
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ori end of clone
252115 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL188841
VERSION AL188841.1 GI:7826945
KEYWORDS GSS; genome survey sequence.

SOURCE
ORGANISM Tetraodon nigroviridis
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthopterygii; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetraodon.

REFERENCE
1 Roest Crollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Fzames, C., Winkler, P., Brotter, P., Quetier, F., Saurin, W. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)

JOURNAL MEDLINE
20296633
10935645

TITLE
Roest Crollius, H., Jallion, O., Dasilva, C., Ozouf-Costez, C., Fzames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetraodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)

JOURNAL MEDLINE
20359837
10899143

REFERENCE
3 (bases 1 to 774)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr)
- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/tetraodon>.

FEATURES
source
1..774
/organism="Tetraodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="252115"
/clone_11b="G"
/note="Genoscope sequence ID : CONG252AE08SP1-end : PUC-Or1"

ORIGIN
Query Match 7.9%; Score 35.8; DB 29; Length 774;
Best Local Similarity 48.7%; Pred. No. 67;
Matches 97; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

46 GGTAGCGCTGCGAGGTGTTCTCAGTACGCGCGCGGCAACACGCGTGGC 105
445 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 386
106 GGTAAATATAGCGGCCCAATCTGAGCTGAACATTTACAGTACGTCGCTAACTCT 165
385 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 326
166 GCACCTGCTCTGCAACATGATGATGATGATGATGATGATGATGATGATGATGAT 225
325 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 266
226 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 244
265 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 247

RESULT 13
BE572585 935 bp mRNA linear EST 15-AUG-2000
LOCUS BE572585
DEFINITION 601325541F1 NCI_CGAP_Mame Mus musculus cDNA clone IMAGE:3603347 5',
mRNA sequence.
ACCESSION BE572585
VERSION BE572585.1 GI:9816305

KEYWORDS
EST.
Mus musculus (house mouse)

SOURCE
ORGANISM Mus musculus
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
1 (bases 1 to 935)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

JOURNAL
Contact: Robert Strausberg, Ph.D.
Email: cga@b-remail.nih.gov
Tissue Procurement: Jeffrey Green M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (ILNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/ILNL at: <http://image.llnl.gov>
Plate: LHM8792 row: a column: 12
High quality sequence stop: 679.

FEATURES
source
1..935
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:3603347"
/sex="female, virgin"
/tissue_type="infiltrating ductal carcinoma"
/dev_stage="5 months"
/lab_host="MDH0B"
/clone_11b="NCI_CGAP_Mame6"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Library constructed by Life Technologies. Investigator providing samples: Jeffrey Green, M.D., NIH"

ORIGIN
Query Match 7.9%; Score 35.8; DB 10; Length 935;
Best Local Similarity 53.1%; Pred. No. 73;
Matches 76; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

104 GCGGAATATATAGGCGCCCAATCTGAGCTGAACATTACAGTACGCGTGGCTACT 163
485 GCTGAAGAAGATCTGGCCAAAGACATCATCTGATACATCTGGAGACTTTGGAAAGC 544
164 CTGCACTTCTCTGCAACTGATGCCGCTAACTGACTGACTTAATCCAGCATGCGC 223
545 CTTTGCTTCTCTTGGCCAAAGGATGACCGTTCTCAGGACTTGAATGAATCAAGTTGG 604
224 GCGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246
605 GTGATACAGATGCCAGGCTTGT 627

RESULT 14
BE604165 435 bp mRNA linear EST 21-AUG-2000
LOCUS BE604165
DEFINITION WBI1413-1416_113 1132S wheat drought stressed leaf cDNA library
Triticum aestivum cDNA clone WBI1413-1416_113_113, mRNA sequence.
ACCESSION BE604165
VERSION BE604165.1 GI:9861435
KEYWORDS EST.
Triticum aestivum (bread wheat)
Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
1 (bases 1 to 435)
Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P.S., Hsia, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen H.T., Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.
The structure and function of the expressed portion of the wheat

JOURNAL
COMMENT

genomes - Drought stressed leaf cDNA library
Unpublished (2000)
Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818

FEATURES

Email: oandersn@pw.usda.gov
Sequence have been trimmed to remove vector sequence and low quality sequence with phred score less than 20
Seq primer: Stratagene SX primer.

```

1..435
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="TAM W101"
/db_xref="taxon:4565"
/clone="WBH1413-416_113_113"
/tissue_type="leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SOLR"
/clone_1fb="Wheat drought stressed leaf cDNA library"
/notes="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site 1: EcoRI; Site 2: XhoI; Plants were given a gradual
stress down to 65% and 78% PWC at Texas Tech University
(D. Zhang in HT Nguyen lab). Total RNA and poly(A) RNA
clones were in vivo excised to give plus-script phagemids
in the TJ Close lab (Choi, Close), at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

```

ORIGIN

Query Match 7.8%; Score 35.6; DB 10; Length 435;
Best Local Similarity 50.0%; Pred. No. 57;
Matches 89; Conservative 0; Mismatches 89; Indels 0; Gaps 0

QY	134	TGAACATTATACACAGTACCGGTGGCGGTAATCTGCACTTCGTCTGCAAACTATATGCCCTGA	193
Db	10	TCAATTGGAGGTCTATGGAAATGGTGAAGATTCAATTGAAGTGCATTCAGCTGCCA	69
QY	194	ACTCTGACTTACTATTAACCGACATGGCGCGCGGTAATGGTGCACAATGTGTGACGGCT	253
Db	70	AGCTGGATTGATTTTAATCTTCATTAAGGCCAGATCATGCAATGATTTCTTTCAATG	129
QY	254	CAGATGACAGCTCAATGCATCTGACCCCAAGTGGCTTGGTAAACAGCGCTACTCTTGA	311
Db	130	GGTCAACAGTTTTTTGTAAACCCGACCAATTAAGTATGTCCTCTGACCGCCCACTGTGTA	187

RESULT 15
BE604467
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

BE604467 498 bp .mRNA linear EST 21-AUG-2000
WHE1413-1416_B07 B07S Wheat drought stressed leaf cDNA library
Triticum aestivum cDNA clone WHE1413-1416_B07, mRNA sequence.

Triticum aestivum (bread wheat)
Triticum aestivum
Triticum aestivum
Eukaryota, Viridiplantae, Streptophyta, Embryophyta, Tracheophyta,
Spermatophyta, Magnoliophyta, Liliopsida, Poales, Poaceae,
Pooidae, Triticeae, Triticum.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 498)	Anderson, O.D., Chao, S., Choi, D.W., Close, T.J., Han, P. S., Hsieh, C.C., Kang, Y., Lazo, G.R., Miller, R., Nguyen, H.T., Rausch, C.J., Seaton, C.L., Tong, J.C. and Zhang, D.	The structure and function of the expressed portion of the wheat genomes - Drought stressed leaf cDNA library	Unpublished (2000)	Contact: Olin Anderson

FEATURES
SOURCE

US Department of Agriculture, Agricultural Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105595773
Fax: 5105595818
Email: oanderson@w.usda.gov
Sequence have been primed to remove vector sequence and low
quality sequence with pyrid score less than 20
Seq primer: StrataGene's SK primer.

1.49

organism="Triticum aestivum"
/mol_type="mRNA"
/cuiotar="TAM M101"
/db_xref="taxon:4565"
/clone="WHB413-1416_B07_B07"
/tissue_type="leaf"
/dev_stage="Full tillering stage"
/lab_host="E. coli SGR"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI. Plants were given a gradual
stress down to 65% and 78% RWC at Texas Tech University
(D. Zhang in HR Nguyen lab). Total RNA and poly(A)⁺ RNA
were prepared, a cDNA library was made, and the cDNA
clones were in vivo excised to give pluscript phagemids
in the T7 close lab (Choi, Close), at the University of
California, Riverside. Plasmid DNA preparations and DNA
sequencing were performed in the OD Anderson lab (all
other authors)."

ORIGIN

Query Match	7.8%	Score 35.6;	DB 10;	Length 498;
Best Local Similarity	50.0%;	Pred. No. 61;		
Matches	89;	Conservative	0;	Mismatches 89;
			Indels	0;
			Gaps	0;

Qy	134	TCGAACATTAAACAGTACCGGTGGCGGTAACTCTGCACCTTGTCTCTCAAACTATGATCCCGTA	193
Db	10	TCAAATTGGACGCTTATGGTGAATGGGTGAAGATTGAGTGAAGTGCATACAGCTCCAAAGA	69
Qy	194	ACTCTGACCTTGCCTATTATCCACGACATGGCGCGGTAAAGTGGACAGATGGTGTGATGGGCT	255
Db	70	AGGTGAGTTTGAATCTTAACTTCCATTAAAGCCGAGATTCATGCAAGATGATCTTTCATG	128
Qy	254	CAAGTACACGCTCAATCGATCTGACCCCAAGTGGCTTGGTAAACAGGCGTACTTCTTGA	311
Db	130	GGTACACAGGTTTTTGTAAACCCGAGCATTTAGTATGTCTCTTGGACCGCCACCTGGCTTA	187

Search completed: March 16, 2004, 04:28:45
Job time : 2237.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456

Sequence: 1 atgaataacaattgttatt.....ctcgctgcacacacgttaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmb1:*

- 1: gb_ba:*
- 2: gb_hcg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sbs:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vt:*
- 15: gb_da:*
- 16: gb_fun:*
- 17: gb_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_ph:*
- 24: em_pl:*
- 25: em_ro:*
- 26: em_sbs:*
- 27: em_sy:*
- 28: em_un:*
- 29: em_vt:*
- 30: em_hcg_hum:*
- 31: em_hcg_inv:*
- 32: em_hcg_other:*
- 33: em_hcg_mus:*
- 34: em_hcg_pln:*
- 35: em_hcg_rsd:*
- 36: em_hcg_mam:*
- 37: em_hcg_vrt:*
- 38: em_ey:*
- 39: em_hgo_hum:*
- 40: em_hgo_mus:*
- 41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	4680	1	ECCSGABDG	X90754 E.coli csgG
2	456	100.0	15046	1	AE000205	AE000205 Escherichia
3	456	100.0	15047	1	D90741	D90741 Escherichia
4	454.4	99.6	1711	1	AF275733	AF275733 Escherichia
5	454.4	99.6	10190	1	AE005315	AE005315 Escherichia
6	454.4	99.6	327773	1	AP002554	AP002554 Escherichia
7	451.2	98.9	10370	1	AE015131	AE015131 Shigella
8	451.2	98.9	306358	1	AE016759	AE016759 Escherichia
9	449.6	98.6	292504	1	AE016981	AE016981 Shigella
10	356.8	78.2	2889	1	CSP515700	CSP515700 Citrobacter
11	332.8	73.0	2067	1	SRU43280	SRU43280 Salmonella
12	332.8	73.0	5103	1	STRJ2301	STRJ2301 Salmonella
13	332.8	73.0	22411	1	AE008749	AE008749 Salmonella
14	331.2	72.6	254050	1	AL627269	AL627269 Salmonella
15	331.2	72.6	301983	1	AE016840	AE016840 Salmonella
16	323	70.8	2920	1	CFR515701	CFR515701 Citrobacter
17	275	60.3	1212	1	BCO131756	BCO131756 Escherichia
18	266.4	58.4	1048	1	STRGFA	STRGFA Salmonella
19	258.2	56.6	2883	1	ESAS15702	ESAS15702 Enterobacter
20	51.2	11.2	301214	1	AE016786	AE016786 Pseudomonas
21	49.4	10.8	246435	2	AC096956	AC096956 Rattus no
22	47.6	10.4	253061	3	AE014834	AE014834 Plasmodium
23	43.4	9.5	234090	2	AC111458	AC111458 Rattus no
24	43.2	9.5	182148	2	AC119000	AC119000 Rattus no
25	43	9.4	3732	3	PFAS27	PFAS27 Plasmodium
26	42.6	9.3	210554	2	AC133982	AC133982 Rattus no
27	41.4	9.1	10709	1	AE015532	AE015532 Shewanella
28	40.4	8.9	172742	2	AC136466	AC136466 Rattus no
29	40.2	8.8	242510	2	AC128507	AC128507 Rattus no
30	39.8	8.7	160460	2	AC024374	AC024374 Homo sapi
31	39.8	8.7	161536	2	AC024318	AC024318 Homo sapi
32	39.8	8.7	179693	9	AL591856	AL591856 Human DNA
33	39.8	8.7	192016	2	AC026540	AC026540 Homo sapi
34	39.6	8.7	7218	6	I66494	I66494 Sequence 14
35	39.6	8.7	257757	3	AE014837	AE014837 Plasmodium
36	39.2	8.6	74524	9	AL512508	AL512508 Human DNA
37	39.2	8.6	190950	2	AL359639	AL359639 Homo sapi
38	38.6	8.5	50820	1	AE014163	AE014163 Streptococcus
39	38.6	8.5	310850	1	AP005142	AP005142 Streptococcus
40	38.2	8.4	4860	6	AR320671	AR320671 Sequence
41	38.2	8.4	186401	2	AC097009	AC097009 Pan trogl
42	38.2	8.4	242730	2	AC099408	AC099408 Pan trogl
43	37.8	8.3	2000	6	AK655393	AK655393 Sequence
44	37.8	8.3	31521	3	CEY47H10A	CEY47H10A Caenorhab
45	37.8	8.3	141489	10	AL591376	AL591376 Mouse DNA

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
ECCSGABDG	ECCSGABDG	E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.	X90754	X90754.1	GI:1147558	csgG gene; csgB gene; csgD gene; csgF gene; csgG gene; csgG gene;	Escherichia coli	1	Hammar, M., Arnyvist, A., Bian, Z., Oleen, A. and Normark, S.	Expression of two csg operons is required for production of
						orfC gene.	Escherichia coli			
						Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;	Escherichia coli			
						Enterobacteriaceae; Escherichia.	Escherichia coli			

flporecitrin- and congo red-binding curli polymers in *Escherichia coli* K-12
 JOURNAL Mol. Microbiol. 18 (4), 661-670 (1995)
 MEDLINE 96414468
 PUBMED 8617489
 REFERENCE 2 (bases 1 to 4680)
 AUTHORS Hammar, M.
 JOURNAL Direct Submission
 Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
 Microbiology and Tumoriology Center, Box 280, S-171 77 Stockholm,
 SWEDEN

FEATURES
 source Location/Qualifiers
 1. 4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:562"
 /map_23.15 minutes
 1. 3140
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141.. 4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="W3110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement (133.. 966)
 /gene="csaG"
 complement (133.. 966)
 /gene="csaG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CAA62277.1"
 /db_xref="GI:1147559"
 /db_xref="GOA:P52103"
 /db_xref="SWISS-PROT:P52103"
 /translation="MQRLLVAWLLSGCTAPPKKARPTLMPRAQSYKDI THLPA
 PTKKLFVSVYNIQDETQFQKYPANFSTAVPOSATMLVTALKDSKMFPIERQGL
 NNLNBRKTIIRAQENGTVAINNRIPIQSTTANIMVGSITIGESNVKSGVGARVFG
 IGADYQYLDQIAVNLRVNVTGSEITISVTSKTIITISVQGVFRFIDYQRLLEGE
 VGYTSNEPVMCLMSAIEITGVIFILINDIGRLMDLONKARONDILVKKRHSVPPE
 S"
 complement (993.. 1409)
 /gene="csaB"
 complement (993.. 1409)
 /gene="csaB"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CAA62278.1"
 /db_xref="GI:1147560"
 /db_xref="SWISS-PROT:P52104"
 /translation="MRVKAIVLMLISPLWAGMTTQFRPNRPGNPNNGAFLNS
 AQAQNSYKQPSYNDDEIETPSALDNFQAIQSGILSINTGKRGKRVNTDYIV
 DIANRQGLQIANTDKRTGOTSTIQVSGIQQNSTDF"
 complement (1434.. 1823)
 /gene="csaE"
 complement (1434.. 1823)
 /gene="csaE"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 production"
 /protein_id="CAA62279.1"
 /db_xref="GI:1147561"
 /db_xref="SWISS-PROT:P52105"

ORIGIN

gene
 CDS
 /translation="MKRYLRWTVAAEFLFAAGNLHAAVEVEPGLTDHTVSSIGHDFY
 PARSKMSDYTGMLTINERSARWGSMITITVQDVIFQTFPLPKRDEKTVAVAL
 IOTEBALNRQINQALSTGDLADEF"
 complement (1828.. 2478)
 /gene="csaD"
 complement (1828.. 2478)
 /gene="csaD"
 /note="necessary for transcription of the csaBA operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CAA62280.1"
 /db_xref="GI:1147562"
 /db_xref="GOA:P52106"
 /db_xref="SWISS-PROT:P52106"
 /translation="MFNEVSHIGHLTLITSSIQATALLQHLKQSLATIGKLANIQ
 RSLDDISSGSIILDMEMADKLIHWQDTLSRKNNTKILLTPEDYPIRDIENMP
 HINGVPSWEDQRRVNGLOVLRQECYFQTLASGLYITHSGNRYNSTESALLTHRE
 KETLNLKIRIGASNNETARSLFISENTVTHLYNLPKLIAYKNRQAVSMANDNLR"
 complement (2626)
 /note="transcriptional start site for csaDfc operon"
 3140
 /note="transcriptional start site for csaBA operon"
 3233.. 3688
 /gene="csaB"
 3233.. 3688
 /gene="csaB"
 /codon_start=1
 /transl_table=11
 /product="nucleation component of curlin monomers"
 /protein_id="CAA62281.1"
 /db_xref="GI:1147563"
 /db_xref="GOA:P39828"
 /db_xref="SWISS-PROT:P39828"
 /translation="MKKKLFLFMMLTILGAPGIAAAGYDLANSEYFAVNEISKSFN
 OAATIGQAGTNSQLRQGGSKLAVAGSGSNRAKITQDTGVNLAIVYIDQGSANDA
 SISGAGVNTAMTIOKSGNKNITGYQKTAIVQRRSQMAIRVTOG"
 3729.. 4184
 /gene="csaA"
 3729.. 4184
 /gene="csaA"
 /note="major subunit"
 /codon_start=1
 /transl_table=11
 /product="curlin/CsaA protein"
 /protein_id="CAA62282.1"
 /db_xref="GI:1147564"
 /db_xref="GOA:P28307"
 /db_xref="SWISS-PROT:P28307"
 /translation="MKLKVAKIAIVFSGSALAGVVPQYGGGNGHGGGNGSGPSE
 LIDQWKNKSEMTVKQFGGNGAAVDQTSNSNVVTQVGFQNNATIAQY"
 3729.. 3788
 /gene="csaG"
 4243.. 4575
 /gene="orfC"
 4243.. 4575
 /gene="orfC"
 /codon_start=1
 /transl_table=11
 /product="putative curli production protein"
 /protein_id="CAA62283.1"
 /db_xref="GI:1147565"
 /db_xref="SWISS-PROT:P52107"
 /translation="MNTLILIAASSQITFTNTTQGDVYTIPEVLTQSCLCRVQIL
 SLRSGSGSQOTQKQKTLSPANQPIALTGLSINTSPDRVXIVTVVSDGSLHLSQO
 MPSSRSK"

sig_peptide
 gene
 CDS

Query Match 100.0%; Score 456; DB 1; Length 4680;
 Best Local Similarity 100.0%; Pred. No. 2,7e-116;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTAAACAATCTAGTGGTGGCTGGATGGAGCC 60
Db 3233 ATGAAAAACAATTGTTATTTATGATGTTAAACAATCTAGTGGTGGCTGGATGGAGCC 3292
Qy 61 GCAGCAGGTTATGATTTAGCTAATTCAGAAATTAACCTTGGCGGTAAATGAATGAGTAAG 120
Db 3293 GCAGCAGGTTATGATTTAGCTAATTCAGAAATTAACCTTGGCGGTAAATGAATGAGTAAG 3352
Qy 121 TCTTCAATTAATCAGCAGCCCAATATGCTCAACCTGGAGCTAATATATGCTCACTTA 180
Db 3353 TCTTCAATTAATCAGCAGCCCAATATGCTCAACCTGGAGCTAATATATGCTCACTTA 3412
Qy 181 CGCAGCAGGAGCTCAAAACTTTGGCGGTTGTTGGCAGAAAGTAGTAGCAACCGGCA 240
Db 3413 CGCAGCAGGAGCTCAAAACTTTGGCGGTTGTTGGCAGAAAGTAGTAGCAACCGGCA 3472
Qy 241 AAGATTGACACAGACAGAGATTAATTAACCTTGATATATTTGATCAGCGCGGACGTCCAC 300
Db 3473 AAGATTGACACAGACAGAGATTAATTAACCTTGATATATTTGATCAGCGCGGACGTCCAC 3532
Qy 301 GATGCGACTATTTCCCAAGTGTCTTATGCTATCTGCGATGATATATCCAAAGCTTCT 360
Db 3533 GATGCGACTATTTCCCAAGTGTCTTATGCTATCTGCGATGATATATCCAAAGCTTCT 3592
Qy 361 GGTATTAAGCAATATATTACACAGATGCTCAAAAAACGGCAATTGATGACAGAGA 420
Db 3593 GGTATTAAGCAATATATTACACAGATGCTCAAAAAACGGCAATTGATGACAGAGA 3652
Qy 421 CAGTGGCAATGCTATTGCGCGTGAACACACCTTAA 456
Db 3653 CAGTGGCAATGCTATTGCGCGTGAACACACCTTAA 3688

```

```

RESULT 2
AE000205
LOCUS Escherichia coli K12 MG1655 section 95 of 400 of the complete
DEFINITION genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
KEYWORDS Escherichia coli K12
SOURCE Escherichia coli K12
ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 10346)
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
TITLE The complete genome sequence of Escherichia coli K-12
JOURNAL Science 277 (5331), 1453-1474 (1997)
MEDLINE 97426617
PUBMED 9278503
REFERENCE 2 (bases 1 to 10346)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 3 (bases 1 to 10346)
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE 4 (bases 1 to 10346)
AUTHORS Plunkett, G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

```

COMMENT

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markborov@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). ** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

repeat_region

promoter

promoter

gene

CDS

```

Location/Qualifiers
1..10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1..48)
/feature="IS3"
<1..5
/feature="factor Sigma70; predicted +1 start at 1094669"
5..32
/feature="factor Sigma70; predicted +1 start at 1094716"
69..392
/feature="b1028"
69..392
/feature="b1028"
/feature="b1028"
/function="orf; Unknown"
/feature="o107; This 107 aa ORF is 27 pct identical (0 gaps)
to 58 residues of an approx. 1888 aa protein FAS2_CANAL
SM: P43098
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAC74113.1"
/db_xref="GI:1787266"
/transl_table="MEDGOVAMRDYLAGLDIASDEHGVLMNTAKSEVAPSIKENLS
LHLTHYKWTESLDPTPLNMPVAFLEISNKPVYGNMLNDPAVDQQRVHRCRAFSFD
AKYVK"
275..299
/feature="b1028"
/feature="factor Sigma32; predicted +1 start at 1094983"
389..1375
/feature="ycdu"
/feature="ycdu"
/feature="synonym: b1029"
389..1375
/feature="ycdu"
/feature="ycdu"
/function="orf; Unknown"
/feature="o328; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYC8B
SM: P47562"
/codon_start=1
/transl_table=11
/product="orf; hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"

```

/translation="MIDYLFIRFODKRLIYVAILGIFRYKNAAGTTPSEDI
GVSGITLVLYNFIPLDKAWKCYTKNDISWPKKONKIELEFLOPVAFLS
LIMLSAMSWGILYQLPSLVALFLISLIGVILFRLMRTISYKQVIAVAKYK
SLTRYVLSVISTVNLSTVNLISPLRNSDSYTBGMULTKSIALLIGVLAINF
FLRSEKRYAFGLRFLQELIDFSSERNASTFPKPLMLRFLILVIEVMITLVSVL
ATLVEMIRFPAVFLCYVCLTYFFYCRFLMNDPMACDMYFRMGHPK"

protein_bind
/gene="ycdu"
/note="central position to predicted promoter: -104"
/bound_moiety="Deor predicted site"
1253. .1280
/gene="ycdu"
/note="No predicted promoter"
/bound_moiety="Deor predicted site"
1337. .1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. .1745
/gene="b1030"
1494. .1745
/gene="b1030"
/function="orf; Unknown"
/note="orf; Unknown"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGSGRCVGYTKKKVLSYELFFKYGEGGIDSLRSPCGOPAHG
VHGSLMSPVDPBRSGLLIPVCAIHEKKARFVRL"

gene
/gene="ycdv"
/note="synonym: b1031"
1513. .1926
/gene="ycdv"
/function="putative structure; Not classified"
/note="o137; This 137 aa ORF is 27 pct identical (1 gap)
to 177 residues of an approx. 80 aa protein RS37_YEAST SW:
P05759"
/codon_start=1
/transl_table=1
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="WCNIRKSPYERTSSINAAVRGILTRFAPAGSPULIAFTVCP
GROLSTPGRGSPRCVQYKKKPVLSYELFFKYGEGGIDSLRSPCGOPAHGVL
SMVSPVPEGRSGLLIPVCAIHEKKARFVRL"

gene
/gene="sexX"
/note="synonym: b1032"
complement (2111. .2198)
/gene="sexX"
/product="tRNA-Ser"
/note="anticodon: GGA; CG Site No. 17878"
/function="tRNA; tRNA"

tRNA
/note="anticodon: GGA; CG Site No. 17878"
2269. .2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338. .2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355. .2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. .3370
/gene="ycdm"
/note="synonym: b1033"
2393. .3370
/gene="ycdm"
/function="putative enzyme; Not classified"
/note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPRA_CORGL
SW: P45637"
/codon_start=1

gene
/gene="ycdm"
/function="putative enzyme; Not classified"
/note="o325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPRA_CORGL
SW: P45637"
/codon_start=1

CDS

Query Match
Best Local Similarity 100.0%; Score 456; DB 1; Length 10346;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAAAAACAATTTGATTATTAATGATTAACATATCTGGCTGGGATTGCAGCC 60
8497 ATGAAAAACAATTTGATTATTAATGATTAACATATCTGGCTGGGATTGCAGCC 60

61 GCAGCGGTTATGATTAGCTAATTCAGATATTAATCTGGGTTAATGATTAAGTAG 120
8557 GCAGCGGTTATGATTAGCTAATTCAGATATTAATCTGGGTTAATGATTAAGTAG 120

121 TCTTCATTTAATCAGGCGCATTAATGCTACAGCTGGGCTAATTAATGATTAAGTAG 180
8617 TCTTCATTTAATCAGGCGCATTAATGCTACAGCTGGGCTAATTAATGATTAAGTAG 180

181 CGGAGGAGGCTCAAACTTTGGCGTTGTCGCAAGAGTAGTAGCAACCGGGCA 240
8677 CGGAGGAGGCTCAAACTTTGGCGTTGTCGCAAGAGTAGTAGCAACCGGGCA 240

241 AAGATTGACGACAGAGATTAATACCTTGCTAATTAATGATTAAGTAGTAGCAAC 300
8737 AAGATTGACGACAGAGATTAATACCTTGCTAATTAATGATTAAGTAGTAGCAAC 300

301 GATGCGATTTTTCGCAAGGCTTAATGTAATGATTAATGATTAATGATTAATGATTA 360
8797 GATGCGATTTTTCGCAAGGCTTAATGTAATGATTAATGATTAATGATTAATGATTA 360

361 GGTAAATAAGCAATATTAACAGATAGTACTCAAAAAAGCAATTTGATGACAGAGA 420
8857 GGTAAATAAGCAATATTAACAGATAGTACTCAAAAAAGCAATTTGATGACAGAGA 420

421 CAGTGGCAATGCTTTTGGCTGACACACAGCTTAA 456
8917 CAGTGGCAATGCTTTTGGCTGACACACAGCTTAA 456

RESULT 3
D90741
LOCUS D90741 15047 bp DNA linear BCT 25-DEC-2002
DEFINITION Escherichia coli K12 genomic DNA. (23.7 - 24.0 min).
ACCESSION D90741 AB001340
VERSION D90741.1 GI:1651509
KEYWORDS Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
csgA; ycdB; cIs; nov; mdoc.
SOURCE Escherichia coli K12
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1
REFERENCE
AUTHORS Oshima, T., Aiba, H., Baba, T., Fujita, K., Hayashi, K., Honjo, A.,
Ikemoto, K., Inada, T., Itoh, T., Kajihara, M., Kanai, K., Kashimoto, K.,
Kimura, S., Kitagawa, M., Makino, K., Masuda, S., Miki, T.,
Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishio, Y.,
Nishio, Y., Saito, N., Sampei, G., Seki, Y., Tagami, H., Takemoto, K.,
Wada, C., Yamamoto, Y., Yano, M. and Horiiuchi, T.

TITLE A 718-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 12.7-28.0 min region on the linkage map

JOURNAL DNA Res. 3 (3), 137-155 (1996)

MEBLINE 97061202

PUBMED 8905232

REFERENCE 2

AUTHORS Ikema, H., Baba, T., Fujita, K., Hayaashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasei, H., Kaishimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.

TITLE The systematic sequencing of the *Escherichia coli* genome in Japan

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 15047)

AUTHORS Mori, H.

JOURNAL Direct Submission

COMMENT Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info., 8916-5 Takayama, Ikoma, Nara 630-01, Japan (E-mail:hmori@gc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)

COMMENT Collaboration Information:

Project:

Group: The Japan E. coli genome DNA sequencing project

Members: (1995.4 - 1996.3)

Alba, H., Baba, T., Fujita, K., Hayaashi, K., Honjo, A., Horiuchi, T., Ikemoto, K., Inada, T., Isono, K., Isono, S., Itoh, T., Kanai, K., Kasei, H., Kaishimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Mikita, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nishimoto, H., Nishio, Y., Oshima, T., Saito, N., Sempel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y. and Yano, M.

Headed by:

Name: Takashi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishior@nibb.ac.jp

Information operator:

Name: Hirotsada Mori

Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan

E-mail: hmori@gc.aist-nara.ac.jp

URL:

The Japan E. coli genome database

http://bwa3.aist-nara.ac.jp.

Location/Qualifiers

1. 15047

/organism="Escherichia coli K12"

/mol_type="genomic DNA"

/strain="K12"

/db_xref="taxon:8333"

/map="23.7-24.0 min"

/clone="Kohara clone #231"

/note="Nucleotide position 1097101-1112147 from the initiation site of Theta (0 min.). This clone is from Kohara lambda miniset library."

<1. 276

/notes="unnamed protein product; ORF_ID:0230#5"

/codon_start=1

/transl_table=11

/protein_id="BAA35822.1"

/db_xref="GI:4062605"

/translation="DLFSSSENALSTFAKPLMLRLFILVIEVMMITLVSLATLVE WRWFEAYFLICVPCLIYFFYCRFLMHNDFMMACDMYFRMGHPNK"

1152 2090

/note="ORF_ID:0231#1

similar to SwissProt Accession Number P45637"

/codon_start=1

/transl_table=11

CDS

gene

/product="Hypothetical 33.0 kd protein in prob-proa intergenic region."

/protein_id="BAA35823.1"

/db_xref="GI:4062606"

/translation="MDIIFHPTPTDQWMEALRKALPOARVANKSGDNDADYAL WHPPVEMLAGDLKAVFALGAGVDSISLKLQAHPEMLPSVPLFLEPTGMQEQEY AVSOVLHWFRRFDYRIQONSSHQDPEYHREDPTIGILAGVGSVAOSLQWRF PLRCKSRKRSKRGVQSPAGREBELSAPLSQCRVLINLPTNPTETGIIINQULELPD GAYLNLARGVHVVEDDLAALDSGKAVMLDVFNRRLPEPSFLMKOHPRTITPHV AAIRPEAPVAIEIRITIAQLERGSVCQVDPRAGY"

2145. 2882

/note="unnamed protein product; ORF_ID:0231#2"

/codon_start=1

/transl_table=11

/protein_id="BAA35824.1"

/db_xref="GI:4062607"

/translation="MYPVLDLMTVASTHAYSTLSDYIAQAKQKIKLPAITDHPDM EDAPHMHFIMNRIMPRVDSGLRGIEANIKNVDSIGSKMFDLSLIIAGFHE PVAPPHDKATNTQMIATIASGNVHIISHPNPKYKIDKAVAAAKGVALEINIS SPISRRGSNDCEVAAAYVDAGWVALGSDSHITAPFMGSFEELKILDAVDPEPER ILNVSPLRLNLFLESKGNAPFAEFADL"

2906. 3460

/note="ORF_ID:0231#3

similar to PIR Accession Number F64035"

/codon_start=1

/transl_table=11

/product="Hypothetical protein H11543"

/protein_id="BAA35825.1"

/db_xref="GI:4062608"

/translation="MNEFSILCRVLGSLYRPODPLVPLFTLIREGLAANPLDQ DELTTRLOKSCDMQVADYVALPIGECAPPRSAWBEATASVAPASEGMPL ADPRAELCTILLAAWSLEDOSTDESEALTELSVILLPCGAFLEGVEMHATTPFW RTMAPLTRDAISAMWDELDEBDEE"

3514. 4053

/note="unnamed protein product; ORF_ID:0231#4"

/codon_start=1

/transl_table=11

/protein_id="BAA35826.1"

/db_xref="GI:4062609"

/translation="WCSISFMAAFSAIRGMNILLSTAITTGIISGIVGWAVALGLL SWAGFLCTAYFACPOGGLKGLISAATLISGVVMAVLIYGSALAPLELITGVYITG IVAFLMCIQAKOLLISFVPGTFIGACATFAGCGDKVLPLSLALGLIFGVAMKSGLM LAARSATFAHREDEIKNKA"

complement(4117. 4950)

/gene="cs9c"

complement(4117. 4950)

/gene="cs9c"

/note="ORF_ID:0231#5

similar to SwissProt Accession Number P52103"

/codon_start=1

/transl_table=11

/product="Curli production assembly/transport component CsgG precursor."

/protein_id="BAA35827.1"

/db_xref="GI:4062610"

/translation="MORLFLIVAVMLISGCLTAPKRAARPTLMPRAQSYDITLPA PTGKITVSVINIDETSCQFKPYPASNSTAVPQATMLATLADSRFRITLREOGLQ NLNLERKIRIAQENGTVAINNRIPLOSLTANNTVGSIIIGVSNVSGGVAKYFG IGADTOYLOLDIIVANLRVNVNSTEILSVNTSKTIIISYQADVFRRIDVORLBGR VGYTSNBPVMLCLMSAIEYGVIFLINDIGRLWDLQNKAKERQNDILVTKYHMSVPPR S"

complement(4977. 5393)

/gene="csgF"

complement(4977. 5393)

/gene="csgF"

/note="ORF_ID:0231#6

similar to SwissProt Accession Number P52104"

/codon_start=1

/transl_table=11

/product="Curli production assembly/transport component CsgF precursor."

/protein_id="BAA35828.1"

/db_xref="GI:4062611"

CDS

gene
CDS
/translation="MRVKAHVLLMLISPLSWAGTMTPOFRNPEGNNNGAFLINS
AOQNSYKDSBYNDPFGIETPSALNDPFOIAOSIIGLISNITGKPGRMVNDYIV
DIANDDQLOLVNDRKGTSTIOVGLONNSTDF"
complement (5418..5807)
/gene="csgB"
complement (5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
CsgB precursor."
/protein_id="BAJ5829.1"
/db_xref="GI:4062612"
/translation="MRVRYRMVAAEFFPAAGN.HAVEVEPGLTDHTVSSIGHDFY
RAPSDESDYGNLTINRPSRWSWITTVQDVIFQTFPLPRDEKIVRAL
IOHEALNRQINQALSTGDLAHDEF"
complement (5812..6462)
/gene="csgD"
complement (5812..6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable csgAB operon transcriptional regulatory
protein."
/protein_id="BAJ5830.1"
/db_xref="GI:4062613"
/translation="MPEVHSIHGHTLLITSSLOATALLQHLKOSAIATKLNIG
RSLDDISSGIIIDMWEADKRIHIMQDTLSRKNNIKILANTPEYPRDLENIG
HINQVFSMEDORRYVNGLOGVKRGCEFTOKLASYLIHSGNRYSTESALLTHRE
KEILNKLRIGASNNELARSLFISENTVKHLVLPFKIAVKRRTQAVSMANDINR"
7217..7672
/gene="csgB"
7217..7672

gene
CDS
Query Match
Best Local Similarity 100.0%; Score 456; DB 1; Length 15047;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTATTTATGATTTTAAACAATATCTGGCGCCCGGATTGCAGCC 60
DB 7217 ATGAAAAACAATTTATTTATGATTTTAAACAATATCTGGCGCCCGGATTGCAGCC 60
QY 61 GCAGCAGTTATGATTTAGCTAATTCAGATATTAATCTCGCGTAAATGATTGAGTAAG 120
DB 7277 GCAGCAGTTATGATTTAGCTAATTCAGATATTAATCTCGCGTAAATGATTGAGTAAG 120
QY 121 TCTTCATTTAATCAGGACCAATATTTGCTCAAGCTGGGACTAATAATAGTCTCAGTTA 180
DB 7337 TCTTCATTTAATCAGGACCAATATTTGCTCAAGCTGGGACTAATAATAGTCTCAGTTA 180
QY 181 CGGCGAGGAGCTCAAACTTTTGGCGTTGTGGCGAAGAGTATGCAACCGGGCA 240
DB 7397 CGGCGAGGAGCTCAAACTTTTGGCGTTGTGGCGAAGAGTATGCAACCGGGCA 240
QY 241 AAGATTGACGACAGAGATATTAACCTTGCAATATTTGATCAGCGCGGACGTGCCAAC 300
DB 7457 AAGATTGACGACAGAGATATTAACCTTGCAATATTTGATCAGCGCGGACGTGCCAAC 300
QY 301 GATGCCGATATTTGCAAGGTGCTTATGTTAATACCTGCGATATTTATCCAGAAAGTTCT 360
DB 7517 GATGCCGATATTTGCAAGGTGCTTATGTTAATACCTGCGATATTTATCCAGAAAGTTCT 360
QY 361 GGTATTAAGCAAAATTTATACAGTATGTTACTCAAAAAACGCAATTTGATGTCAGAGA 420
DB 7577 GGTATTAAGCAAAATTTATACAGTATGTTACTCAAAAAACGCAATTTGATGTCAGAGA 420
QY 421 CAGTGCAGATGCTATTTGCGCTGACACAGCTTAA 456
DB 7637 CAGTGCAGATGCTATTTGCGCTGACACAGCTTAA 456
7637 CAGTGCAGATGCTATTTGCGCTGACACAGCTTAA 7672

RESULT 4
AF275733
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Escherichia coli
Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 1711)
Ulrich,G.A., Keen,J.E. and Elder,R.O.
Mutations in the csgD promoter associated with variations in curli
expression in certain strains of Escherichia coli O157:H7
Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
AF275733.1 GI:14039399

REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
2 (bases 1 to 1711)
Ulrich,G.A., Keen,J.E. and Elder,R.O.
Direct Submission
Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal
Research Center, State Spur 18D, Clay Center, NE 68933, USA
1..1711
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="43895 Red Variant"
/serotype="O157:H7"
/db_xref="taxon:562"
757..1212
/gene="csgB"
757..1212
/codon_start=1
/transl_table=1
/product="CsgB protein"
/protein_id="BAK53211.1"
/db_xref="GI:14039400"
/translation="MKNKLFWMLTILGAPGIAAAGVDLANSENPANVELSGSSN
QAALIGAGTNNASQIQRQGSKLAVVDESSNRAKIDQGDYNAITIDQAGSANDA
SISQAGYGNATMIQSGSNKANITQYGTQATVAVQROSMALRVQR"
1253..1711
/gene="csgA"
1253..1711
/codon_start=1
/transl_table=1
/product="CsgA protein"
/protein_id="BAK53212.1"
/db_xref="GI:14039401"
/translation="MKLRYAATATATVFGSSALAGVFPQYGGGNGHGGGNNSGPNS
ELNIYQGGNSALATQADARNSDLTITQGGSGNADVQGGSDSSIDITQGFQNSA
TLQDNGDSIMTYKQKRGSGNAAVQDTASNSTVNTQVGFNNATAHQY"

ORIGIN

Query Match
Best Local Similarity 99.6%; Score 454.4; DB 1; Length 1711;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTATTTATGATTTTAAACAATATCTGGCGCCCGGATTGCAGCC 60
DB 757 ATGAAAAACAATTTATTTATGATTTTAAACAATATCTGGCGCCCGGATTGCAGCC 60
QY 61 GCAGCAGTTATGATTTAGCTAATTCAGATATTAATCTCGCGTAAATGATTGAGTAAG 120
DB 817 GCAGCAGTTATGATTTAGCTAATTCAGATATTAATCTCGCGTAAATGATTGAGTAAG 120
QY 121 TCTTCATTTAATCAGGACCAATATTTGCTCAAGCTGGGACTAATAATAGTCTCAGTTA 180

Db 877 TCTTCATTTAATCAGCAGCACCCATAATGTGTAAGCTGGGAGTAATATATAGTCTCAGTTA 936
Qy 181 CGCAGGAGAGCTGCTAACTTTTGGCGGTGTGGCGAAGAGTAGTAGCAACCGGCA 240
Db 937 CGGCGGAGAGCTGCTAACTTTTGGCGGTGTGGCGAAGAGTAGTAGCAACCGGCA 996
Qy 241 AAGATTGACGACAGAGAGATATATACCTTGATATATGATCAGGCGGCGAGTCCAC 300
Db 997 AAGATTGACGACAGAGAGATATATACCTTGATATATGATCAGGCGGCGAGTCCAC 1056
Qy 301 GATCCAGATATTTGCGAAGGTGCTTATGTAATATGATGATATATCCAGAAAGTTCT 360
Db 1057 GATCCAGATATTTGCGAAGGTGCTTATGTAATATGATGATATATCCAGAAAGTTCT 1116
Qy 361 GGTATATAAGCAAAATATACACATGTACTCAAAAACGGCAATGTAGTCAGAGA 420
Db 1117 GGTATATAAGCAAAATATACACATGTACTCAAAAACGGCAATGTAGTCAGAGA 1176
Qy 421 CAGTCGCAATGGCTATTCGGGTGACACACAGCTTAA 456
Db 1177 CAGTCGCAATGGCTATTCGGGTGACACACAGCTTAA 1212

RESULT 5
AE005315 10190 bp DNA linear BCT 21-MAR-2001
LOCUS Escherichia coli O157:H7 EDL933 genome, contig 1 of 3, section 139
DEFINITION
ACCESSION AE005315 AE005174
VERSION AE005315.1 GI:12514572
KEYWORDS
SOURCE
ORGANISM Escherichia coli O157:H7 EDL933
Escherichia coli O157:H7 EDL933
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS 1 (bases 1 to 10190)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
TITLE
JOURNAL
MEDLINE 21074935
PUBMED 11206551
REFERENCE 2 (bases 1 to 10190)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glaesner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Weich,R.A. and Blattner,F.R.
Direct Submission
TITLE
JOURNAL Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
SOURCE
1. 10190
/organism="Escherichia coli O157:H7 EDL933"
/mol_type="genomic DNA"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:155864"
/note="enterohaemorrhagic"
617. 1072
/gene="csdB"
/note="synonym: Z1675"
617. 1072
/gene="csgB"
/function="structural component; Cell exterior
constituents: Surface structures"
/note="Residues 1 to 151 of 151 are 100.00 pct identical
to residues 1 to 151 of 151 from Escherichia coli K-12

gene
CDS
11206551

Strain MG1655: B1041"
/codon_start=1
/transl_table=11
/product="minor curlin subunit precursor, similar to CsgA"
/protein_id="AAG55787.1"
/db_xref="GI:12514573"
/translation="MKKKLLFMMLTITGAPGIAAAGVDLANSEYPAVNELSKSPN
OATIGQGTNLSQILROGSKLLAVAAQSSSRPAKIDQGDYINLAYIDQAGSANDA
SISGAYGNTMMIIQKSGNKNITQYGTQKTAIVQKQSMALRYQR"
1113. 1571
/gene="csgA"
/note="synonym: Z1676"
1113. 1571
/gene="csgA"
/function="regulator; Cell exterior constituents: Surface
structures"
/note="Residues 1 to 152 of 152 are 96.71 pct identical to
residues 1 to 151 of 151 from Escherichia coli K-12 Strain
MG1655: B1042"
/codon_start=1
/transl_table=11
/product="curlin major subunit, coiled surface structures;
cryptic"
/protein_id="AAG55788.1"
/db_xref="GI:12514574"
/translation="MKLLKVAATATATPSSGALAGVTPYCGGCGGHHGGGNGSPNS
ELNITQYGGNSALQADANSDLTITQHGCGADVGQSDSDSIDTQRGNSA
TLIDPMNGKSDHMTVKQFGGNGAIVDQASNTVNVQVGFNNATAHQY"
1630. 1962
/gene="csgC"
/note="synonym: Z1677"
1630. 1962
/gene="csgC"
/function="putative structure; Not classified"
/note="Residues 1 to 110 of 110 are 98.18 pct identical to
residues 1 to 110 of 110 from Escherichia coli K-12 Strain
MG1655: B1043"
/codon_start=1
/transl_table=11
/product="putative curli production protein"
/protein_id="AAG55789.1"
/db_xref="GI:12514575"
/translation="MNALLALALSSQITENTQOGDMYTIIPVTLTQSLCQVOLL
SLRSGSGQGTQKQKTLSPANQPIALTKLSINISPDVKYIVTVSDGSLHLSQO
WPSSSEKS"
2083. 2394
/gene="Z1678"
2083. 2394
/gene="Z1678"
/function="orf; Unknown function"
/note="Residues 1 to 103 of 103 are 98.05 pct identical to
residues 1 to 103 of 103 from Escherichia coli K-12 Strain
MG1655: B1044"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG55790.1"
/db_xref="GI:12514576"
/translation="MFRPFLNSLMLGSLFFPFLIAGSTAQGVHFGQIVBPAQDV
STQSPVEMNCPONGSVGKTVSSKALMSNVNQAQASVYQVLDKQKLAWNIEY
N"
2489. 3022
/gene="Z1679"
2489. 3022
/gene="Z1679"
/function="orf; Unknown function"
/note="Residues 1 to 177 of 177 are 100.00 pct identical
to residues 1 to 177 of 177 from Escherichia coli K-12
Strain MG1655: B1045"
/codon_start=1
/transl_table=11
/product="putative polyprotein"
/protein_id="AAG55791.1"

gene
CDS
2489. 3022


```
JOURNAL      Syet. Appl. Microbiol. 23 (3), 315-324 (2000)
MEDLINE      20557356
PUBMED       11108008
REFERENCE    3
AUTHORS      Yokoyama, K., Makino, K., Kubota, Y., Matanabe, M., Kimura, S.,
              Yutsudo, C. H., Kurokawa, K., Ishii, K., Hattori, M., Abe, H., Iida, T.,
              Yamamoto, K., Hayashi, T., Yasunaga, T., Honda, T., Sasaki, C. and
              Shinagawa, H.
TITLE        Complete nucleotide sequence of the prophage VP1-Sakai carrying the
              Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
              O157:H7 strain derived from the Sakai outbreak
JOURNAL      Gene 258 (1-2), 127-139 (2000)
MEDLINE      20564182
PUBMED       1111050
REFERENCE    4
AUTHORS      Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K.,
              Yokoyama, K., Han, C.-G., Ohtsubo, E., Nakayama, K., Murata, T.,
              Tanaka, M., Tobe, T., Iida, T., Takami, H., Honda, T., Sasaki, C.,
              Ogasawara, N., Yasunaga, T., Kuhara, S., Shiba, T., Hattori, M. and
              Shinagawa, H.
TITLE        Complete genome sequence of enterohemorrhagic Escherichia coli
              O157:H7 and genomic comparison with a laboratory strain K-12
JOURNAL      DNA Res. 8 (1), 11-22 (2001)
MEDLINE      21156231
PUBMED       11258796
REFERENCE    5
AUTHORS      (bases 1 to 327773)
              Ohnishi, M., Kurokawa, K., Makino, K., Yasunaga, T., Shinagawa, H. and
              Hayashi, T.
TITLE        Direct Submission
              Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
              Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
              Japan (E-mail: ken@gen-info.osaka-u.ac.jp)
              URL: http://www.gen-info.osaka-u.ac.jp/, Tel: 81-6-6879-8365,
              Fax: 81-6-6879-2047)
              genome project.
COMMENT      Location/Qualifiers
FEATURES     1..327773
SOURCE       /organism="Escherichia coli O157:H7"
              /mol_type="genomic DNA"
              /strain="O157:H7"
              /sub_strain="RIMD 0509952"
              /db_xref="taxon:83334"
              150..2258
              /gene="ECe1032"
              150..2258
              /note="similar to YCBY ECOLI gi|1787181 percent identity
              99 in 702 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /evidence=not experimental
              /transl_table=1
              /product="putative oxidoreductase"
              /protein_id="BAB34455.1"
              /db_xref="GI:13360492"
              /translation="MNSLFASTARGELRLKTELNLGAVRCOVVGGVHFKGPNRLV
              YQSIWMSRLASRIWLPGECKVYSDLDLYLQVAINTEMNRCATPRVHSGNDTIV
              RNSOYGAKVKDAIVDAFTKRLPREVNDKADPIRVWVLHKESTALDLSGDHL
              LRGRDRAGIAPIKETTLAAIVMSRGMQPTPLDPMCGSTLLIEAMLTADAPGL
              HRGWRSGMAOHEALIWQEVAAEQTPARKGLABSSHFYGSDSDARVIGTARAPL
              LAGIGELTFEVDVAULTNPLPKPEYGVLSNPYGERLDSBPALILHSLRLIMK
              NQFGAMNLSLFSAPDLSCLQLRADKQYKANGPLDVCVGNVVAESTPSKRAMA
              EDVANRLRKTKKEKARROGICRYRLYDADLERYVAVNDRYADWVYVGYAPKTI
              DAHAKRQLPITIAITISVGIAPNKALTKRKRGKNOYQKLGKEFLVETVNA
              HUNVNLDTYDGTGLPDRIRIARMLQMSKADPLNLSYGSATVHAGLGASTTT
              VDMGRYILEMAERMLRLNGLTGRARHLQADCLMLEANEOPDLIFIDPPTFENSKR
              MEDAFVQDRDIALMKDLKRLIRAGTITFNNKRGFRMDGLAKLGLKAOETOKT
              LSGDFARROIHNCWLITAA"
              2270..4177
              /gene="ECe1033"
              2270..4177
              /note="similar to YCBY ECOLI gi|1787182 percent identity 99
              in 635 aa (Conserved in E.coli K-12)"
              /db_xref="GI:13360492"
              /protein_id="BAB34456.1"
              /db_xref="GI:13360493"
              /translation="MNSLISHGANLSTSDAPLDLNAELHIEDNERVCLVGRNACKST
              LMKLNRQGLDGRITIIYEDLIVARLQDDPPRVGSGVYDFVAGIEQABYLKRYH
              DISLVNDRPESEKINELAKYQOBDLHNNLWQENRINEVLAQGLDNPVLSLSGQ
              WLRALAGALVSNPRVLLDEPNHLDIETIDWLEGLFTENGRIIFISDRSIRN
              MATRIYDLDRGKLTVPYGNQDYLLKEBSALRVLELNAEFDKRLAEOEYVROKIA
              RPTNERRVRLKAMRBRERGRREVMTGAKQVVEASGKIVEMEVQVYQKOL
              VKDSAOVLRDRIKALIGNCCGKTTLLKMLGQLQADSGRIHVTGKLEVAIFDQRA
              ELDPKTVMDNLAEKQEVWVNGKPRVLYGLDPLFHPKAMTPVRLSGEERNLL
              LARPLFRSNLLIIDEPTNDLSDVTELELIDSLDSYQGVTLVSHDRQFVNTVTEW
              IFESGKIGRVGVDHARGQOEYVALKQPAVKNEERPAKAEYVRSKSLSYKL
              ORELEQIPOLLEDEAKLEALQTVADASFSPQHEQTKVLAADMAAEOLEQAFER
              MEYVELKNGS"
              4307..5560
              /gene="ECe1034"
              4307..5560
              /note="similar to POIA ECOLI gi|1787183 percent identity
              99 in 417 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /evidence=not experimental
              /transl_table=1
              /product="paraguat-inducible protein A"
              /protein_id="BAB34457.1"
              /db_xref="GI:13360494"
              /translation="MCEHHDAKIIILSCQDMVLVPLHEGQAKACRCPCTTITVAV
              DAPQRPAYALALFMLILNSLPFNVMNVAVGTSITLLETIGVFESEYASLGT
              FLFVOLVPAFCILITILLVNRALPRLEKQRLAVFLQKTWMAEFLAGVSVFV
              KLMAYGSIIGVSSFLPWCFLQVQFQCYDRMLMDIDLAPMELQPLRGVTVIR
              QGSRGSCCTAIIIPADEPVCRCGTGKGVRRRNLONTALTLVLSIMLYLPATLPI
              VTDLIGKMPSTIIAGVILMSGSSVVAANYFIASIMVPTLKMILAIAMLCMDAKHG
              KRDSERHILYEVAFYGRMSMIDVPYIVALSALVRMGLSITYPANGALFALVTH
              TMSAMTFDPLSDRQPESEHERS"
              5565..7205
              /gene="ECe1035"
              5565..7205
              /note="similar to PQIB ECOLI gi|1787184 percent identity
              99 in 546 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /evidence=not experimental
              /transl_table=1
              /product="paraguat-inducible protein B"
              /protein_id="BAB34458.1"
              /db_xref="GI:13360495"
              /translation="MESNNGEAKIOKVMNSPWIPIVYALIGAWVLFYHSHQGE
              VTULTANAEGIEGKTTIKSRSDVGVAVESATLADDTLHVEIKRLNSGMEKLAHKT
              VFWVVKQIGREGISGTLISGVYELQGAKGSKMDKYDLDSPLAPDAGIRV
              VLSKKGQGLSPGDPVPRGVRSVSTSPDPOKRYISQVLFNAYDPLVSNVPR
              KQSGIADVDTLSAMFRVKSGLTTLGSGVSPVDESLDQAPAPPTAPLYYDQYS
              IQDLSYDTHLIDYALFFQDSVRGLQGAPEFRGIRLQTSKVPFPAPNMRQTFDDDR
              IIPVIRIRIEPRNLKQLEBNADVHLELKRGLRGLKGNLVYGLAYDLDFYPT
              PATIGIEBPNGYQIIPVSGLAQIOORLEALDKINKLPLNPMIEQATSTLSQST
              KQNLQTLDSMNKTIASQMOQLPTDMOSTRELNRMOGQPOESAAYVKNVAVMDQML
              DOYRLQPLTKLTLESNMLVFEAKKQKQDEPRGAKQ"
              7217..7765
              /gene="ECe1036"
              7217..7765
              /note="similar to YWBA ECOLI gi|1787185 percent identity
              98 in 182 aa (Conserved in E.coli K-12)"
              /codon_start=1
              /evidence=not experimental
              /transl_table=1
              /product="hypothetical protein"
              /protein_id="BAB34459.1"
              /db_xref="GI:13360496"
```



```

/locus_tag="SF1029"
/note="Residues 1 to 184 of 184 are 99 pct identical to
residues 1 to 184 of a 184 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287169.1"
/codon_start=1
/transl_table=1
/product="putative oxidoreductase component"
/protein_id="AAN42653.1"
/db_xref="GI:24051315"
/translation="MNEFSTICRVUGSLYRPODPLVPLEFTLIRECKLANPFEQ
DELTIRLQKSCDMTVQVADYNALFTGDECAVPYKSAVEDATEAVAPLSEKML
ADTPADHIGTLLASMLQDSTEDSEALFTLFESEYLLPMCGAFLGVEAHATPFM
RTMAPLTRDAITLISAMDELEEDSEE"
2811..3350
/gene="ycdZ"
/locus_tag="SF1030"
2811..3350
/gene="ycdZ"
/locus_tag="SF1030"
/note="Residues 1 to 179 of 179 are 98 pct identical to
residues 1 to 179 of a 179 aa protein from Escherichia
coli K12 ref: NP_415554.1"
/codon_start=1
/transl_table=1
/product="orf, conserved hypothetical protein"
/protein_id="AAN42654.1"
/db_xref="GI:24051316"
/translation="MCSISFMAAPSAIRGNMILISAITTIGSLGWAVSLGLL
SWAGFLGTAFACPOGGLKGLAISAAFLSGVWAMVITIGSLAPLLETIGYVITG
IVAFMCIQAKQLILSFVPTFGACATFACQGWKVLPSLALGLVGYAMKNSGLM
LAARSAKTRAHEOEIKNKA"
complement(3414..4189)
/gene="csgG"
/locus_tag="SF1031"
/pseudo
complement(3414..4189)
/gene="csgG"
/locus_tag="SF1031"
/note="Similar to Escherichia coli K12 curli production
assembly-transport component, 2nd curli operon gi: 1787274
(278 aa). BLAST with identity of 97% in 257 aa. This CDS
contains an in-frame stop codon. The sequence has been
checked and is believed to be correct."
/pseudo
/codon_start=1
/transl_table=1
complement(4454..5662)
/locus_tag="SF1032"
complement(4454..5662)
/locus_tag="SF1032"
/note="Residues 1 to 402 of 402 are 100 pct identical to
residues 1 to 402 of a 402 aa protein from Plaemid R100
ref: NP_052925.1"
/codon_start=1
/transl_table=1
/product="transposase of Tn10"
/protein_id="AAN42655.1"
/db_xref="GI:24051317"
/translation="MCELDILHDSLYOCPELHLKRLNSLTLAGHALDCKTLTTEL
GRNLPKARTGHNIRIDRLGNRHLHERLAAYRMHASFICSGNTWPIVAVDSIR
EOKRLVLRASVALHGRSVTLYEKAPLSEOCSSKKAHOPLADLASILPSTPLIYS
DAGRKVPYKSVKLGWYLSRVKQYVADLCAENKPIENLHDMSSSHKTYGYR
LTENPISQCLILYKSRKRNORSRTTHCHPSPKTISAKEPWILATNLVEIR
TPKLVNYSKRMQIEETFDLKSPAVGLGRHSRTSSSEFDMLIALMLQLTCLM
AGVAHQKQGMKHFQANTVRNRNVLSTVRLGMEVLRHSGYITREDLSVAATLLTONL
FTNGYVLGKL"
complement(5995..6384)
/gene="csgE"
/locus_tag="SF1033"
complement(5995..6384)
/gene="csgE"
/locus_tag="SF1033"
/note="Residues 1 to 129 of 129 are 98 pct identical to

```

```

residues 1 to 129 of a 129 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287173.1"
/codon_start=1
/transl_table=1
/product="curli production assembly-transport component,
2nd curli operon"
/protein_id="AAN42656.1"
/db_xref="GI:24051318"
/translation="MKRILRIVAAERLFAAGNLHAVBEVYGLTHTVSSIGHDFY
RAFSDKWSDYTNLITNERSARWGSVITTVQDVIFQFLPPLKRSKSEVFEAL
IOTEALNRQINQALLSTDLDLAEF"
complement(6389..7069)
/gene="csgD"
/locus_tag="SF1034"
complement(6389..7069)
/locus_tag="SF1034"
/note="Residues 1 to 226 of 226 are 100 pct identical to
residues 1 to 216 of a 216 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287174.1"
/codon_start=1
/transl_table=1
/product="putative 2-component transcriptional regulator
for 2nd curli operon"
/protein_id="AAN42657.1"
/db_xref="GI:24051319"
/translation="WRSIKKGFMEVHSIHTLLITKPSLQATALLQHLKQSL
ATIKLNIORSLDDISSGSIILDMNEADKRLHYQDTLSRNNNIKILLTPED
YPRDIDFNHPINQVFAAMEQERVAVNGLOVLEGEYFQTKLASYLTISGNRYNS
TESALATREHEIINKLRIGASNNEIARSLPISNTVYKTHLYNLFKJIAVNRTOVNS
WANDNLR"
7766..8248
/gene="csgB"
/locus_tag="SF1035"
7766..8248
/gene="csgB"
/locus_tag="SF1035"
/note="Residues 10 to 160 of 160 are 99 pct identical to
residues 1 to 151 of a 151 aa protein from Escherichia
coli O157:H7 EDL933 ref: NP_287175.1"
/codon_start=1
/transl_table=1
/product="minor curli subunit precursor, similar to CsgA"
/protein_id="AAN42658.1"
/db_xref="GI:24051320"
/translation="MDQVQGDNNKNTLFLMMLTILGPGIAAGYDLANSBYNFAV
NELSKSPNOAIIIGAGTNNNSQLRGGSKLLAVAOEGSSNPAKIDQTDVNLAYI
DQGSANDASISGAYGNMTMIQKSGNKNKITQYQKTAVVVQROSQWAIHVTOR

```

Query Match 98.9%; Score 451.2; DB 1; Length 10370;
 Best Local Similarity 99.3%; Pred. No. 5.8e-115;
 Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 1 ATGAAAAAATAATGTTATTTATGATGTTAAACATATCTGGGTGGCCCTGGATGCAACC 60
DB 7793 ATGAAAAAATAATGTTATTTATGATGTTAAACATATCTGGGTGGCCCTGGATGCAACC 7852
QY 61 GCAGCAGGTTATGATTTAGTTAGTTCAGATATTAATCTGGCGTAAATGAATTAGTAAG 120
DB 7853 GCAGCAGGTTATGATTTAGTTAGTTCAGATATTAATCTGGCGTAAATGAATTAGTAAG 7912
QY 121 TCTTCATTATTAATCGGACCATTAATGTGTCACTGGAGCTAATTAATAGTCTCAGTTA 180
DB 7913 TCTTCATTATTAATCGGACCATTAATGTGTCACTGGAGCTAATTAATAGTCTCAGTTA 7972
QY 181 CGGACGAGGAGCTCAAAATCTTTGGCGTGTTCGCAAGAAGGTAGTACGACCGGCA 240
DB 7973 CGGACGAGGAGCTCAAAATCTTTGGCGTGTTCGCAAGAAGGTAGTACGACCGGCA 8032
QY 241 AAGATTACACAGACAGAGATTAATCACTTGACATATTAATGATCAGCGGACGATGCCAAC 300
DB 8033 AAGATTACACAGACAGAGATTAATCACTTGACATATTAATGATCAGCGGACGATGCCAAC 8092

```


/product="Putative conserved protein"
/protein_id="AA079731.1"
/db_xref="GI:26107532"
/translation="MPROPDGYREIKOTMOLSLTAKSKPELVQELKGVLLAVNL
VRYQWIKAEHLKGVNQLSFSKCGVMMLMTLQASPGRIPELMRDILASMGOLV
KLPTRRAPFPRVVERP"
7651..8013
/locus_tag="c1258"
7651..8013
/locus_tag="c1258"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AA079732.1"
/db_xref="GI:26107533"
/translation="MLGACFTSMGTSFYTGPPMGAELVLMMPKACMLQSSLYSL
LDGEQVLFSEWLNKKQYEGFEFYQIFSGNIVASCSGVSNSAHYARCTDFMRH
YSLNNYACQLOKROACLS"
7835..8269
/locus_tag="c1259"
7835..8269
/locus_tag="c1259"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AA079733.1"
/db_xref="GI:26107534"
/translation="MKDLSFRPLALLLVPALEPVRHRTGALTLCTVTL
ITLLASGRLVSKCISESIFIRRHIECMAGRTYFTFSLIAGWYRQKKEY
QTLHYENRYVLANVDGLVLSQSYRSKRKYINQHRLL"
8477..8553
/locus_tag="c5534"
8477..8553
/locus_tag="c5534"
/product="rRNA-Arg"
/function="rRNA; rRNA"
complement(8832..9290)
/locus_tag="c1260"
complement(8832..9290)
/locus_tag="c1260"
/function="putative enzyme; Transposases Insertion
Sequence Associated"
note="Escherichia coli O157:H7 ortholog: z4324"
/codon_start=1
/transl_table=1
/product="Putative Transposase"
/protein_id="AA079734.1"
/db_xref="GI:26107535"
/translation="MASAADARIVSHYERMLIEYHKAWKSGTCVESLRMOTRD
LERNVVIAFLAVMLGLRQISSETQNDCKILPTTEWKLIMVKLEGQLPSQTP
TLKAACTLGRWDSKRTGRPMVMMGMFRLDMVGYEVMSLDEI"
complement(9302..9667)
/locus_tag="c1261"
complement(9302..9667)
/locus_tag="c1261"
/codon_start=1
/transl_table=1
/product="Hypothetical protein"
/protein_id="AA079735.1"
/db_xref="GI:26107536"
/translation="MAERLGRITQKRVITVCREADIMVLYLHYKXSHGCRFVVRTRQNS
98.9%; Score 451.2; DB 1; Length 306358;
Best Local Similarity 99.3%; Pred. No. 5.5e-115;
Matches 453; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTCTTATTTATGATGTTAAACAATCTGGTGGCCCTGGGATTGCCAGCC 60
DB 44773 ATGAAAAACAATTCTTATTTATGATGTTAAACAATCTGGTGGCCCTGGGATTGCCAGCC 44832
QY 61 GCACAGAGTTATGATTTAGTAAATTCAGATATACTCGGGTAAATGAATTAGTAG 120
DB 44833 GCACAGAGTTATGATTTAGTAAATTCAGATATACTCGGGTAAATGAATTAGTAG 44892
QY 121 TCTTCATTTAATCAAGCAGCCATTAATGGTCAAGCTGGAGCTAATATAGTGCTCAGTTA 180
DB 44893 TCTTCATTTAATCAAGCAGCCATTAATGGTCAAGCTGGAGCTAATATAGTGCTCAGTTA 44952
QY 181 CGCAGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACGACCGGGCA 240
DB 44953 CGCAGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACGACCGGGCA 45012
QY 241 AAGATTGACGACAGGAGATTATACCTTGATATTTAGTACAGCGGCGAGTCCAAC 300
DB 45013 AAGATTGACGACAGGAGATTATACCTTGATATTTAGTACAGCGGCGAGTCCAAC 45072
QY 301 GATGCCAGTATTTGCGAAGGTGCTTATGTAATACCTGGATGATTTATCCAGAAAGTTCT 360
DB 45073 GATGCCAGTATTTGCGAAGGTGCTTATGTAATACCTGGATGATTTATCCAGAAAGTTCT 45132
QY 361 GGTATTAAGCAATATTTACACAGTAGTCTCAAAAAGCGCAATTTGTGACAGGA 420
DB 45133 GGTATTAAGCAATATTTACACAGTAGTCTCAAAAAGCGCAATTTGTGACAGGA 45192
QY 421 CAGTCGCAATGCGCTAATTCGCGTACACACAGTTAA 456
DB 45193 CAGTCGCAATGCGCTAATTCGCGTACACACAGTTAA 45228

RESULT 9
LOCUS AE016981
DEFINITION Shigella flexneri 2a str. 2457T section 4 of 16 of the complete genome.
ACCESSION AE016981
VERSION AE016981.1
KEYWORDS GI:30040616
SOURCE
ORANISM Shigella flexneri 2a str. 2457T
Shigella flexneri 2a str. 2457T
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Shigella.
REFERENCE
AUTHORS Wei, J., Goldberg, M.B., Burland, V., Venkatesan, M.M., Deng, M., Fournier, G., Mayhew, G.F., Plunkett, G. III, Rose, D.J., Darling, A., Mau, B., Perna, N.T., Payne, S.M., Runyen-Janecky, L.J., Zhou, S., Schwartz, D.C., and Blattner, F.R.
Complete Genome Sequence and Comparative Genomics of Shigella flexneri Serotype 2a Strain 2457T Infect. Immun. 71 (5), 2775-2786 (2003)
12704152
2 (bases 1 to 292504)
TITLE JOURNAL PUBMED
AUTHORS
REFERENCE
TITLES
JOURNALS
FEATURES
source
1. 292504
/organism="Shigella flexneri 2a str. 2457T"
/mol_type="genomic DNA"
/strain="2457T"
/serotype="2a"
/db_xref="taxon:198215"
complement(223..441)
/gene="infA"
/locus_tag="S0884"

/translation="MSOEYEDKEVTLTKLSGRRLBALLLIVLPAWMLAALLSF
 NSPDSMSOTAMHEPIHNLGGMWGLADLTLEFPGWAAVTPVIVGGCNPAMRHOS
 SDEYIDYFAVSLRITIGVALITLSCGLAIAADITWYASGVIVSLISTLOPLHS
 SGGTIALLCWAAGITLFTGMSWTIARKLGMILNITFTSNRRTDWTWDEDEYE
 DDEEYEDNHHQHSRRRLIRGLARKRLAEKFIIMPGEQTDAAIFSGRMDDE
 EITTYARVGVADPDVPLSGNRAITDPEYDPLTGNSTNITEPVAAATAATOSMA
 APEVETQTPVAVSPQPTVAMQVPGQTEPIYAPABEGYPOOSQAOQVAVQ
 YNEPLQOQVPOQPYAPABEQPAQPYAPABEQVAPABEQVAPABEQVAPABEQ
 TEOTVOQPAOEBLYQOPOVPOQVPEPVEPVEPVEPVEPVEPVEPVEPVEPVE
 OLAMQYPIPEPVEPVEPVEPVEPVEPVEPVEPVEPVEPVEPVEPVEPVEPVE
 TVAAVFSLSGSGRPQVKEGIGPOLRPRIATPRLASVGIKLPSPRAAEKA
 REAONQYSDQVNDDBIDAMQODELROFQTOQOQVGEQYQHDVNVNEDADAA
 EAEILRPAQTOQOQVSGEOPAGANFSLDFEFSMKALLDDGHEPLFTPIYAVQ
 QOQVAVQOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQV
 APOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQVPOQV
 DLTPRESEVPEVTFALBQMARLEAALBRLIADVNTSPCVITREFEILAPGV
 KAARISNLSRLARSLTVAARVAVVEITGKPVGELPNKRQIVYLAIEVDNAFRD

Query Match 98.6%; Score 449.6; DB 1; Length 292504;

Best Local Similarity 99.1%; Pred. No. 1.5e-114; Matches 452; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTAACATATCTGGGTCGCGCTGGGATTGCAGCC 60
 Db 210815 ATGAAAAACAATTGTTATTTATGATGTTAACATATCTGGGTCGCGCTGGGATTGCAGCC 210874
 Qy 61 GCAGCAGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGAGTAAAG 120
 Db 210875 GCAGCAGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGAGTAAAG 210934
 Qy 121 TCTTCATTTAATCAAGCAGCAGCATAATTTGTCAGCTGGGAGACTAATATAGTCTCAGTTA 180
 Db 210935 TCTTCATTTAATCAAGCAGCAGCATAATTTGTCAGCTGGGAGACTAATATAGTCTCAGTTA 210994
 Qy 181 CGGCAAGGAGCTCAAAACTTTTGGCGGTGTTGGCGCAAGAGGTAGTACACCGGGCA 240
 Db 210995 CGGCAAGGAGCTCAAAACTTTTGGCGGTGTTGGCGCAAGAGGTAGTACACCGGGCA 211054
 Qy 241 AAGATTGACAGACAGAGATTTAATTAATCTTGATATTTGATCAGCGCGGCGAGTGGCCAC 300
 Db 211055 AAGATTGACAGACAGAGATTTAATTAATCTTGATATTTGATCAGCGCGGCGAGTGGCCAC 211114
 Qy 301 GATGCGAGTATTTGTCAGAGTGTCTTATGTAATCTGCGATGATTTCCAGAAAGTTCT 360
 Db 211115 GATGCGAGTATTTGTCAGAGTGTCTTATGTAATCTGCGATGATTTCCAGAAAGTTCT 211174
 Qy 361 GGTATATAAGCAATATTATACAGATGATGTACTCAAAAACCGCAATTTGATGTCAGAGA 420
 Db 211175 GGTATATAAGCAATATTATACAGATGATGTACTCAAAAACCGCAATTTGATGTCAGAGA 211234
 Qy 421 CAGTCGCAATGCGTATTCGCGGTGACACACGTTAA 456
 Db 211235 CAGTCGCAATGCGTATTCGCGGTGACACACGTTAA 211270

RESULT 10

CSP515700

2889 bp DNA linear BCT 24-JUN-2003

CSP515700 Cytrobacter sp. Fec2 csbB gene, csbA gene and csbD gene.

Accession AJ515700.1 GI:31790491

Version AJ515700.1 GI:31790491

Keywords csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation

Component of curlin monomers; regulatory protein.

Source Cytrobacter sp. Fec2

Organism Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

Enterobacteriaceae; Cytrobacter.

Reference 1 Zogaj X., Bokranz W., Nimtz M. and Bowling U.

Production of Cellulose and Curli Fimbriae by Members of the Family

Enterobacteriaceae Isolated from the Human Gastrointestinal Tract

Infect. Immun. 72 (7), 4151-4158 (2003)

Journal Reference 2 (bases 1 to 2889)

AUTHORS Romling U.
 TITLE Direct Submission
 JOURNAL Submitted (11-Nov-2002) Romling U., Microbiology and Tumorbiology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES location/Qualifiers
 source 1..2889

/organism="Cytrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 complement(121..862)
 /gene="csbB"
 complement(212..862)
 /gene="csbD"
 /function="necessary for transcription of the csbA
 operon"

/codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="CAD56670.1"
 /db_xref="GI:31790492"
 /translation="MFNBFPSIHGHTLLITKPSLOAPALLOHLKOSLALSGKLNIO
 RSLDIDISSCVIVDMMEDAKLIHYMDNLSKNNLKTLLTPTDTPYPRDLENP
 HINQVYVADDERVNVNGQITLGEYFSKSLAYLTTHSGNTRINSSSALTLHRS
 KEILNKLKIGASNEIARSLFISNTVKTLYNFKIAYKRTQAVSWANDLRR"
 /length="2078"
 /gene="csbB"
 /length="2078"
 /gene="csbB"
 /function="nucleator"
 /codon_start=1
 /transl_table=1
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56671.1"
 /db_xref="GI:31790493"
 /translation="MKRLKFLMLTYGAGRIASATSYDLAHSYRPAVNELSSSFN
 OALITGVGTNNSSAKMDEGSKLSIVSDEGSRARVSDGAVNPAVIAQSGHSDA
 SISQSNYGNAMTIQKSGNKNITQYGTQAVVOROSMAIRVIO"
 /length="2571"
 /gene="csbA"
 /length="2571"
 /gene="csbA"
 /function="curli subunit"
 /codon_start=1
 /transl_table=1
 /product="curlin-csbA protein"
 /protein_id="CAD56672.1"
 /db_xref="GI:31790494"
 /translation="MKLLQVAFALIVSGSALAGVPOWGGGGGGGGSSGPESTL
 SIYSGVNAALALQSDARKSDITIHQNGGNGDVGCGSNGSTIDITQNGFKNNATI
 DQWNGKSDITVSGYGGHNAALVNGTASDSSVLHVQFGNNATANYQ"

ORIGIN

Query Match 78.2%; Score 356.8; DB 1; Length 2889;
 Best Local Similarity 86.4%; Pred. No. 1.3e-88; Matches 394; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTGTTATTTATGATGTTAACATATCTGGGTCGCGCTGGGATTGCAGCC 60
 Db 1623 ATGAAAAACAATTGTTATTTATGATGTTAACATATCTGGGTCGCGCTGGGATTGCAGCC 1682
 Qy 61 GCAGCAGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGAGTAAAG 120
 Db 1683 GCAGCAGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAAATGAATGAGTAAAG 1742
 Qy 121 TCTTCATTTAATCAAGCAGCAGCATAATTTGTCAGCTGGGAGACTAATATAGTCTCAGTTA 180
 Db 1743 TCTTCATTTAATCAAGCAGCAGCATAATTTGTCAGCTGGGAGACTAATATAGTCTCAGTTA 1802
 Qy 181 CGGCAAGGAGCTCAAAACTTTTGGCGGTGTTGGCGCAAGAGGTAGTACACCGGGCA 240

ORGANISM *Salmonella enteritidis*
Salmonella enteritidis
Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales
Enterobacteriaceae; *Salmonella*.
1 (bases 1 to 2067)
AUTHORS Doran, J.L., Collinson, S.K., Burian, T., Santos G., Todd, E.C.,
Murro, C.K., Kay, C.M., Banseer, P.A., Peterkin, P.I. and Kay, W.W.
TITLE DNA-based diagnostic tests for *Salmonella* species targeting *agfA*,
the structural gene for thin aggregative fimbriae
J. Clin. Microbiol. 31 (9) 2263-2273 (1993)
JOURNAL MEDLINE
94013373

REFERENCE
3 (bases 1 to 2067)
AUTHORS
Collinson, S.K., Doran, J.L., Banser, P.A. and Kay, W.W.
TITLE
Direct Substitutions
JOURNAL
Submitted (13-DEC-1995) S. KAYAN 0-11

```

    /organism="Salmonella enteritidis"
    /mol_type="genomic DNA"
    /strain="27655-3b"
    /db_xref="taxon:592"
    /map="between putA and pyrC"
    571..576
    588..603
    696..1151
    gene
    -35_signal
    -10_signal
    gene
    CDS
    /gene="agfB"
    696..1151

```

```

/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/translation_table=1
/product="agfB precursor"
/protein_id="AAC43598.1"

```

```

/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial
stem_loop
gene
CDS
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=1
/product="AgfC"
/protein_id="AAC33600.1"
/db_xref="GI:1184715"
/translation="MHRTLLALALSNQITFTTQGDIVTVPQVTLNEPCVQVILIT
SVRQVGGSGHTQCKQTLLPANQPIELRLSVNISSEDSYKLIITVSDGSLHSQ
MPSPAQ"

```

Query Match	73.0%	Score 332.8	DB 1	Length 2067
Best Local Similarity	83.1%	Pred. No. 6,4e-82		
Matches 379	Conservative 0	Mismatches 77	Indels 0	Gaps 0
QY	1	ATGAAAAACAAATGTTATATGATGTTAACAATAC	CTGGGTGCGCTGGATGACGC	60
Db	696	ATGAAAAACAAATGTTATATGATGTTAACAATAC	CTGGGTGCGCTGGATGACGC	60
QY	61	GCACAGCTTATGATTTAGCTATATGAGAAATTA	CTCGCGTAAATGAATTGATAG	120
Db	756	GCACAAATATATGATCTGGCTCGTCAGAATTA	TTTTCGCGTAAATGAATTGACAG	815
QY	121	TCTTCATTTATACGACGACCAATATTTGTC	CAAGCTGGACTAATATATAGTCCAGTTA	180
Db	816	TCTTCATTTATACGACGACCAATATTTGTC	CAAGCTGGACTAATATATAGTCCAGTTA	180
QY	181	CGCAGGAGGCTCAAAATCTTTGGCGTGTGTC	GCAAGAAAGTATGACCAACCGGCA	240
Db	876	CGCAGGAGGATCAAAACTATTTGCCGTTATTT	CAAGAAGGAGAAATTAATTCGGCGC	935
QY	241	AAGATTGACCAACAGAGATTTATTAACCTTG	CAATATTTGATCAGGCGGCACTGCGAAC	300
Db	936	AAGTCGACGACGAGGAAATTTATTAACCTTG	GCAATATTTGATCAGGCGGCACTGCGAAC	300
QY	301	GATGCAAGTATTTGCAAGGTGCTTATGATAT	ACTGCCATGATTTCCAGAAAGGTTCT	360

```

Db      996  GATGCACATATATGCGCAAGCGCTTACGCTATAGTACGACGCTATATCCAGAAAAGCTTCT 1055
Oy      361  GGTATATAAAGCAATATTTACACAGTATGTTACTCAAAAACCGCAATTGTAGTCAGAGA 420
Db      1056  GGAATATAGGCGCAATATTACCCAGTACCGTACGCAAGAAACAGCAGTGTAGTCAGAAA 1115
Oy      421  CAGTCGCAATGGCGCTATTGGGTGACACAGCTTAA 456
Db      1116  CAGTCGCAATGGCTATTGGCTGCTACCCAGCCCTAA 1151

RESULT 12
STAJ2301 5103 bp DNA linear BCT 15-NOV-2000
LOCUS Salmomella typhimurium cs9g, cs9f, cs9g, cs9d, cs9b, cs9a, and k
DEFINITION
ACCESSION AJ002301 GI:2739232
VERSION AJ002301.1 GI:2739232
KEYWORDS cs9a gene; cs9b gene; cs9c gene; cs9d gene; cs9e gene; cs9f gene;
SOURCE Salmomella typhimurium
ORGANISM Salmomella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmomella.
REFERENCE 1
AUTHORS Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.
TITLE Culi fibers are highly conserved between Salmomella typhimurium
and Escherichia coli with respect to operon structure and
regulation
JOURNAL J. Bacteriol. 180 (3), 722-731 (1998)
MEDLINE 98117058
PUBMED 9457880
REFERENCE 2 (bases 1 to 5103)
AUTHORS Romling, U.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
source
1..5103
Location/Qualifiers
1..5103
/organism="Salmomella typhimurium"
/mol_type="genomic DNA"
/strain="SR-11"
/db_xref="taxon:602"
<1..322
/note="ORF179a"
/codon_start=2
/transl_table=1
/product="hypothetical protein"
/protein_id="CAA05311.1"
/db_xref="GI:2739233"
/db_xref="GOA:O54290"
/db_xref="SWISS-PROT:O54290"
/translacion="CTLSGWMALVLIHSGALAPHEIVSYVLTVGIAPLMCIQAKO
LLLSFVPGTFIGACATFAGQGMRLVLPGLGLIFGVAMKNSGLMLASRRQHSANT
AVTK"
complement(361..1194)
/gene="cs9g"
complement(361..1194)
/gene="cs9g"
/function="lipoprotein"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cs9g protein"
/protein_id="CAA05312.1"
/db_xref="GI:2739234"
/db_xref="GOA:O54291"
/db_xref="SWISS-PROT:O54291"
/translacion="MPRLILVAALVLSGCCITAPPKQAKFTLMPRAQSYKDLTHLPA
PTGKIFSVSVNIQDETGQFKPYPSNSFTAPQASATMLVTALKDSRMFIPLEQGLQ
NLNBRKIRAAQENGTAVMNNRIPLQSLTPANIMVSGSIIGESNVKSGGVARFYG
IGADTOYQLDQIAVNLRVNVSTGEILSSVTSKTIISYEVOAGVPRFIDYQRLLEG

```

```

gene
complement(1221..1637)
/gene="cs9f"
complement(1221..1637)
/gene="cs9f"
/codon_start=1
/transl_table=1
/product="Cs9f protein"
/protein_id="CAA05313.1"
/db_xref="GI:2739235"
/db_xref="SWISS-PROT:O54292"
/translacion="MRVKAHVLIILMFSPLTWAGNMTFOPRNPVRGNGPNNGSFLLNS
AQONSKYKDPAYDNDRGIERPSALDNFTQAIQSIIIGGLTLNITGPRGAVTWDFTI
DIARDQQLQUNVDRKTRSTIEVGLQSTDF"
complement(1664..2059)
/gene="cs9b"
complement(1664..2059)
/gene="cs9b"
/codon_start=1
/transl_table=1
/product="Cs9b protein"
/protein_id="CAA05314.1"
/db_xref="GI:2739236"
/db_xref="SWISS-PROT:O54293"
/translacion="MKRYLTIVVAELLPATGNLHANEVEVEPGLTDHRTVSIGHE
FYRAFSDKMSSEYGNLTINERPSARGSWITTVNODVIFQTFLPKQDFEKTIVF
ALAQTEALRLRRQIDQTLSTDLARDF"
complement(2064..2714)
/gene="cs9d"
complement(2064..2714)
/gene="cs9d"
/function="transcriptional regulator"
/codon_start=1
/transl_table=1
/product="Cs9d protein"
/protein_id="CAA05315.1"
/db_xref="GI:2739237"
/db_xref="GOA:O54294"
/db_xref="SWISS-PROT:O54294"
/translacion="MFNEVHSHGHITLLITKPSIQATLALQHLKQSLATNGKLANIQ
RSLEIDISAGCIVLMDMEADKLIHYQDNLSRKNNIKTLLNTPDQYPRRIENMP
HINGVFATKADSDQEHVSGLOGILRGECYFQOKLASIYITSGNVRVNSTSALLTHE
KEILNKIRIGASNNRIRSRSPISENTVTKTHLYLFKLAVKRTQAVNSMNDINLR"
complement(2869..2896)
/gene="cs9DERG"
complement(2869..2896)
/gene="cs9DERG"
complement(2891..2896)
/gene="cs9DERG"
3344..3376
/gene="cs9BA"
3344..3349
/gene="cs9BA"
3371..3376
/gene="cs9BA"
3469..3924
/gene="cs9B"
3469..3924
/gene="cs9B"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=1
/product="Cs9B protein"
/protein_id="CAA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translacion="MKRKLFLPMULTIIGAGCIATPATNYDLARBEYNFANVLSKSSNS
QAAIIGGVGNDNARVROEGSKLSVTSQSGNNRAVDAGNNVNFAYIICOTGNANDA
SISQSAAGNSAAAIQKSGKKNANITQVGTQTAVVVQKQSHMAIRVQR"
3469..3531
sig_peptide

```

```

mat_peptide      /gene="csaB"
                  3532..3921
/gene="csaB"
/product="CsaB protein"
/function="nucleator"
3966..4421
/gene="csaB"
3966..4421
/function="fiber subunit"
/codon_start=1
/evidence-experimental
/transl_table=11
/product="CsaB protein"
/protein_id="CA05117.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/translacion="MKLKVAFALIVSGSALAGVVPOMGGGNGGNSGSDST
LSITQGSANALALQSDARKSETTQSGVNGADYCGGADNSTTLETONGFRNAT
IDQWNAKNSDITVGQYGGNNAALVNQTSBDSVWVRGVGNATANY"
3966..4025
/gene="csaA"
4026..4418
/gene="csaA"
/product="CsaA protein"
/function="fiber subunit"
4483..4809
/gene="csaC"
4483..4809
/gene="csaC"
/codon_start=1
/transl_table=11
/product="CsaC protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translacion="MTLLALLASNOTFTTTOGGDIYTVIPQVTLNEPCVCQVLT
SVRDVGQGSHTQOKTSLPLNQPILSLRSLVNSSEDSVKITIVSDQSLHLSQO
WPSAQ"
4483..4533
/gene="csaC"
4534..4806
/gene="csaC"
/product="CsaC"
4840..5103
/gene="ORF103"
4940..55103
/gene="ORF103"
/codon_start=1
/transl_table=11
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translacion="MSVKKKNIPIALIGCTCAPIFHSAGVQGVTHFGAIVEPL
CDISTHAENI"

ORIGIN
Query Match      73.0%; Score 332.8; DB 1; Length 5103;
Best Local Similarity 83.1%; Pred. No. 6.3e-82;
Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATGAAAACAAATTTTATTTATGATTTACATATCTGGGCGCCCTGGGATTCGACCC 60
DB 3469 ATGAAAACAAATTTTATTTATGATTTACATATCTGGGCGCCCTGGGATTCGACCC 60
QY 61 GCAGCAGGTATGATTTAGTAACTTCAATTAATCTTCCGGTAAATGAATGATGAAG 120
DB 3529 GCAGCAGGTATGATTTAGTAACTTCAATTAATCTTCCGGTAAATGAATGATGAAG 120
QY 121 TCTTCAATTAATCGCAGCAGCATAATTTGGTCAAGTGGGACTAATATATGCTCAGTTA 180
DB 3589 TCTTCAATTAATCGCAGCAGCATAATTTGGTCAAGTGGGACTAATATATGCTCAGTTA 180

```

```

QY 181 CGCAGCAGCAGCTCAAACTTTTGGCGGTGTTGCGCAGAGAGTAGTACCAACCGGCA 240
DB 3649 CGCAGCAGCAGCTCAAACTTTTGGCGGTGTTGCGCAGAGAGTAGTACCAACCGGCA 240
QY 241 AAGATTGACACAGCAGCAGGATTTATACCTTGATATTTGATCAGCGGCGAGTCCAC 300
DB 3709 AAGATTGACACAGCAGCAGGATTTATACCTTGATATTTGATCAGCGGCGAGTCCAC 300
QY 301 GATGCGGATTTTGGCAGAGTCTTATGATATAGTCCATGATTAATCCAGAAAGTCT 360
DB 3769 GATGCGGATTTTGGCAGAGTCTTATGATATAGTCCATGATTAATCCAGAAAGTCT 360
QY 3829 GGAATATGAGCAGATTTATCCAGTAGGAGTACGAGAAACAGCAGTTGTAGTGAGAA 420
DB 421 CAGTCGCAATATGCTTTTGGCGTACACACAGCTTAA 456
QY 3889 CAGTCGCAATATGCTTTTGGCGTACACACAGCTTAA 456
DB 3924 CAGTCGCAATATGCTTTTGGCGTACACACAGCTTAA 456

RESULT 13
AE008749 22411 bp DNA linear BCT 23-Apr-2003
LOCUS Salmonella typhimurium LT2, section 53 of 220 of the complete
ACCESSION AE008749 AE006468
VERSION AE008749.1 GI:16419641
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2
Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 22411)
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Lettelle, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
Stonking, T., Nhan, M., Waterston, R., and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE 2 (bases 1 to 22411)
AUTHORS
CONSTRM
TITLE
JOURNAL
The Salmonella typhimurium Genome Sequencing Project
Direct Submission
Submitted (29-MAR-2001) Genome Sequencing Center, Department of
Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI43283

COMMENT
Coding sequences below are predicted from manually evaluated
computer analyses, using similarity information and the programs:
GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
GeneMark; http://opal.biology.gatech.edu/GeneMark/
EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
http://ecocyc.PangeaSystems.com/ecocyc/
The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
http://kinich.cifn.unam.mx:8850/db/regulondb\_intro.frameset
This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such

```

FEATURES

SOURCE

as compressions and repeats; all regions were covered by sequence from more than one m13 subclone.

Location/Qualifiers

```
1..22411
  /organism="Salmonella typhimurium LT2"
  /mol_type="genomic DNA"
  /strain="LT2; SGSC 1412; ATCC 700720"
  /db_xref="ATCC:700720"
  /db_xref="taxon:99287"
  /note="LT2"
  /gene="phoH"
  /gene="phoH"
  /note="synonym: STM1126"
  /gene="phoH"
  /note="putative RBS for phoH; RegulonDB:STMS1H001398"
  /protein_id="AAL20057.1"
  /db_xref="GI:16419642"
  /translation="MCANRQVENAGDYLQKYLISMGKQAVIKARREARVLRSD
  RSHQREBSYVTSVOMGVBAIGMARSDRTSPVKANEAQAHYLAIDSKQIFAT
  YDGLCKGTWISAAKAAEALIHQDVERIIVTRPVLOADELGLPGDIAEKPAFRRPV
  YDLIKRLGASFMOCLEBEIGKVEIAPFAVRGRTPENAVIILDEAQNAAQMKMP
  LTRIGENTVTVNGDITQCCLPRGSRSLDALRFEDEVGVIVHPKDCVSAALC
  ORTHAYS"
  complement(1414..2303)
  /gene="STM1127"
  complement(1414..2295)
  /gene="STM1127"
  /note="similar to E. coli orf, hypothetical protein
  (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
  identity in aa 15 - 264"
  /codon_start=1
  /transl_table=1
  /product="putative transcriptional regulator"
  /protein_id="AAL20058.1"
  /db_xref="GI:16419643"
  /translation="MEPQPRLLKPKGLIDLTGAMQKSLTRASQRIAYIILAPRQVQ
  SSIALDRETOGAGTAVIRFCRTLYGKGFODPKDLAIELATSDSDSPLLDAEVS
  SDDAHAIGLKQNTISNVLSETLLMLDQGVLDLRLHCHSYVIRFGVSSGITALD
  MKHLMRIIGLRGDAVSNHFMVMOATILKAGDAVGVSHGSTSPETVHSLRLAQACA
  TTVAITRNLSGPLCEADFCILNQRQMLQGSISGRKAQLFVFDLLYTLVSSPE
  QARSKLRKTMALDMTK"
  complement(2298..2303)
  /gene="STM1127"
  /note="putative RBS for STM1127; RegulonDB:STMS1H001399"
  complement(2580..4088)
  /gene="STM1128"
  complement(2580..4076)
  /gene="STM1128"
  /note="similar to E. coli putative cotransporter
  (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
  identity in aa 7 - 478"
  /codon_start=1
  /transl_table=1
  /product="putative sodium/glucose cotransporter"
  /protein_id="AAL20059.1"
  /db_xref="GI:16419644"
  /translation="MITHSGEIVLVILFGLIAMLVGVYFSRROKADYPRGGR
  VRGAAVSVPATLSSITFMSIPAKAFSTWMTIIGQYLAIALILPVFFYIFPRK
  LKTSATIELEARDVRCRLPASMSPMLFHIGRAITIFVLARPPAIADIVTLA
  LISVWCIIYTMGGIEGVITWDVILQGLLHSGAIIIFIVCLKVGQSIDELFVTLQDA
  DKFPAQFMSTSTSTPVLMIGLFLFANIQFASQDVVQRYIVDSISETKTLIT
  NAKLAVIIVFFFAISGALFYVYQGHQLLPAGNTGILPLFVTEMPIAGLILIA
```

RBS

gene

CDS

```
AIFPAQSSISSLSNLSISCFNSDIYORLSHKRRTPENRMKILAVILVAGLISSAAS
VILYMADESEIWDAPNLSLIGMGPMKLPMLGIFPRKANGSAVLGIIIVITVYLG
RYATDLNFFFYIGLSLVISGVIFAPLAPAPPLTLDEKPEPVTL"
  complement(4083..4088)
  /gene="STM1128"
  /note="putative RBS for STM1128; RegulonDB:STMS1H001400"
  complement(4413..5099)
  /gene="STM1129"
  complement(4413..5093)
  /gene="STM1129"
  /note="similar to E. coli putative enzyme (AAC76255.1);
  Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
  226"
  /codon_start=1
  /transl_table=1
  /product="putative inner membrane protein"
  /protein_id="AAL20060.1"
  /db_xref="GI:16419645"
  /translation="MSILARLEQSVHENGGLIVSCOPVPGSPMDKPRIVAMAQAAAS
  AGAAYVIEGENTRTYRPHLSVYIIGIRDLTGSFRTYPIYQDDALAQACADIT
  AFDSFRRPVDISLIRLRHGLMADSTVNEGISCHQKIEFGITLSTGTP
  ITPEVPLAVTQLSHAGCRVIAEGRVYTPALANALIEHGAIVTGSATIRIHICQ
  WFSHAVXR"
  complement(5094..5099)
  /gene="STM1129"
  /note="putative RBS for STM1129; RegulonDB:STMS1H001401"
  5399..6759
  /gene="STM1130"
  5599..6759
  /gene="STM1130"
  /note="similar to E. coli orf, hypothetical protein
  (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
  identity in aa 40 - 403"
  /codon_start=1
  /transl_table=1
  /product="putative inner membrane protein"
  /protein_id="AAL20061.1"
  /db_xref="GI:16419646"
  /translation="MGQMKNPFKKMTLMALCSVAITTSQVATLLPDIEPLKNGTG
  AIDNNGVIVYGLTAGTSWKYKIDLKQHKOMERIKSPGARGQSVFNLDELVYVG
  GVSKGKESPLQVYSDVYKSPVNTQKDDTISPVGLTGTVKLBETWVILGVN
  EHIIFDKYFIDIAAADADESKNKVITYVFKPADYFENKIVFPIYNKENTWKNAGEL
  PGAGTAGSSVVMGNPFMLINGELKPLPTDVIYRAMWMDKLTMLNSQLPSPRGQ
  QQESLAGPESGSHCVILVUGGANFPKAKQNTNGKRYSHGCIKKRDEYGLINGH
  WQYGMKQKPLGVGSVSYGDEVFLIGENAKGPVSVISFTWRDNLTK"
  6793..7497
  /gene="STM1131"
  6793..6798
  /gene="STM1131"
  /note="putative RBS for STM1131; RegulonDB:STMS1H001402"
  6805..7497
  /gene="STM1131"
  /note="similar to E. coli orf, hypothetical protein
  (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26%
  identity in aa 4 - 239"
  /codon_start=1
  /transl_table=1
  /product="putative outer membrane protein"
  /protein_id="AAL20062.1"
  /db_xref="GI:16419647"
  /translation="MKIKKYLGMVSIFARSSVLAQATLDRHEVADRTRINKORAI
  IEKLPNGIGPYVAVSKSGVDGEQDKHLSDYANAIELGVSRYRYKTDNPFVLOPGTI
  FESGPDSTIKPYLRQGYNFDSGVTMAGRYRYOYARTKATYASDEKTRNPDYIYGVF
  DELKLEYNFTWMSDDQIKFDNKKTYNEHNVALMKLNSFTPYVEGVGVAVRNNTDER
  QTRYRYGLQYHP"
  7768..9060
  /gene="STM1132"
  7768..7773
  /gene="STM1132"
  /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
  7780..9060
  /gene="STM1132"
  /note="similar to E. coli putative transport protein
```



```
/gene="flg"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AU513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein Flg"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLMTAKTGELDAQNTMDVIANLNANVSTNGFKRQAVFEDL
LYQTRPGAQSSEQTLPGLQIGTVRPAATERLHQGNLSQTNNSKDAIKGOGR
FOVWLPGDTSAYTRDGSFQVNDQGLVAGGFQVQPAITIPANLSTIGDGVSVT
OQGOAIPVOGQNLNLTFTMDTGLSEIGNLYETOSGAPNESPPGANGGLLYQGY
VETSNVNVAEELVIMVIOVRAVEINSKAVSTTDMQLKLTQL"
/complement(7146..7901)
/gene="flg"
/locus_tag="t1742"
/complement(7146..7901)
/gene="flg"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein Flg"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="WDHAIYTMGCAASQTLNQAVTASNLNASTPGFRAQLNLRAV
PNDLSLATRLVYASTPGADMTQGLDYTSRPLDVALQODGMLVQAAADAGEYTRN
GNIQVPTGQLTIQHPVIGEGPTVEGSEITIAADGTISALNPGDPPTVAIVGR
LKLVAEGNEVORSDDGLFRLTAEOERGAVLADPSIRIMSGVLBSNVKPEAMT
DMIANARFEMOMYITTSVDNEGRANQLSMS"
/complement(7922..9133)
/gene="flg"
/locus_tag="t1743"
/complement(7922..9133)
/gene="flg"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein Flg"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="WSESQAVSGLNAAATNLDIVGNINANSATYGFSGTASFPADMPA
GSKVGLGVKAVGITQDFDTGTTNTVGRGLDVAIISQNGFFRLVDNSGVFYSRNGQFKL

Query Match      72.6% Score 311.2; DB 1; Length 301983;
Best Local Similarity 82.9% Pred. No. 1.7e-81;
Matches 378; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTGTTATTTATGATGTTAAACAATACCTGGGTGCGCTGGATTGCAGCC 60
DB 37807 ATGAAAAACAATTGTTATTTATGATGTTGAACAATACCTGGGTGCGCTGGATTGCAGCC 37748
QY 61 GCAGCAGGTTATGATTTAGCTTAATTCAGAAATTAATCTCGCGGTAATGAATTGAGTAA 120
DB 37747 GCACAAATTTATGATCTGCTGCTGCTAGATATATTTTGGCGTAATGAATGAAGCAAG 37688
QY 121 TCTTCATTATTCAGGCAGCCATATTTGTCAGCTGGGACTAATATATAGTCTCAGTTA 180
DB 37687 TCTTCATTATTCAGGCAGCCATATTTGTCAGCTGGGACTAATATAGTCTCAGTTA 37628
QY 181 CGGAGGAGGCTCAAACTTTTGGCGGTGTTGGCAAGAGTAGTAGCAACCGGGCA 240
DB 37627 CGCAGGAGGATCAAACTATGTCCTTATTTCACAAGAGAGAAATATATCGGGCG 37568
QY 241 AAGATTGACCAAGAGATTAATCACTTGCAATATATGATCAAGCGGCGGAGTCCCAAC 300
DB 37567 AAGATGACCAAGAGGAGAAATTAATCACTTGCAATATATGAGCAAAACGGGCAATGCCAAC 37508
QY 301 GATGCCAGTATTTGCAAGGTGCTTAATGTAATACTGCGATGATTAATCCAGAAAGTTCT 360
```

```
DB 37507 GATGCCAGTATATGCAAAAGCGCTTAACGTAATAGTGGGCTATTATTCAGAAAGTTCT 37448
QY 361 GGTATATAAGCAAAATATTATACACAGTATGTAATCAAAAAACGGCAATTGTAGTCAGAGA 420
DB 37447 GGAATATAAGGCATATATTACCAAGTACGGTACGAGAAACAGCAGTTGTATGTCAGAAA 37388
QY 421 CAGTCGCAATGCTATTTGCGGTGACACCAACGTTAA 456
DB 37387 CAGTCGCAATGCTATTTGCGGTGACACCAACGCTAA 37352
```

Search completed: March 15, 2004, 22:49:54
Job time : 1967.17 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaacaaatcgtatc.....ctcgctgacacacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 337363 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002s:*\n7: geneseqn2003as:*\n8: geneseqn2003bs:*\n9: geneseqn2003cs:*\n10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3 AAC64620	AAC64620 Escherich
2	332.8	73.0	456	3 AAC64618	AAC64618 Salmonell
3	100	21.9	100	7 ACD68810	ACD68810 E. coli K
4	100	21.9	100	7 ACD68812	ACD68812 E. coli K
5	100	21.9	100	7 ACD68811	ACD68811 E. coli K
6	45.8	10.0	779	1 AAN40272	AAN40272 Sequence
7	38.2	8.4	4860	8 ADA31934	ADA31934 DNA encod
8	37.8	8.3	2000	7 ADA71938	ADA71938 Rice gene
9	37.4	8.2	456	3 AAC64623	AAC64623 Agfa::PT3
10	36.6	8.0	456	3 AAC64629	AAC64629 Agfa::PT3
11	35.2	7.7	4890	2 AAX34648	AAX34648 Promoter
12	35	7.7	646	2 AAO62647	AAO62647 Fibronect
13	34.6	7.6	361	2 AAQ73066	AAQ73066 Agfa sequ
14	34.6	7.6	361	2 AAT74141	AAT74141 Salmonell
15	34.6	7.6	456	2 AAQ87467	AAQ87467 Agfa sequ
16	34.6	7.6	456	2 AAT74142	AAT74142 Salmonell
17	34.6	7.6	456	3 AAC64622	AAC64622 Agfa::PT3
18	34.6	7.6	456	3 AAC64617	AAC64617 Salmonell
19	34.6	7.6	456	3 AAC64625	AAC64625 Agfa::PT3
20	34.6	7.6	456	3 AAC64624	AAC64624 Agfa::PT3
21	34.4	7.5	4455	7 ABE17054	ABE17054 Human MP2
22	34.4	7.5	5184	9 ADE25651	ADE25651 Human CDN
23	34.4	7.5	6325	4 AAK80005	AAK80005 Human imm

C	24	34.4	7.5	9555	7 ABX34542	ABX34542 Human mdd
C	25	34.2	7.5	1083	5 AAS76745	AAS76745 DNA encod
C	26	33.4	7.5	12426	4 AAS46482	AAS46482 Tumour su
C	27	33.8	7.4	1149	3 AAB65947	AAB65947 E. coli p
C	28	33.8	7.4	1149	7 ACA18571	ACA18571 Prokaryot
C	29	33.8	7.4	2655	6 ABE70427	ABE70427 Streptoco
C	30	33.8	7.4	2655	6 ABE70427	ABE70427 Streptoco
C	31	33.8	7.4	6108	4 ABL08478	ABL08478 Drosophila
C	32	33.8	7.4	6808	4 ABL08488	ABL08488 Drosophila
C	33	33.8	7.4	110000	6 ABE71527_06	Continuation (7 of
C	34	33.6	7.4	606	3 AAC43625	AAC43625 Arabidops
C	35	33.6	7.4	1299	7 ACA48239	ACA48239 Prokaryot
C	36	33.2	7.3	1371	7 ACA29389	ACA29389 Prokaryot
C	37	33.2	7.3	4590	5 AAB24065	AAB24065 Yeast AOD
C	38	33	7.2	1269	7 ACA48121	ACA48121 Prokaryot
C	39	33	7.2	1837	7 ACA47731	ACA47731 Prokaryot
C	40	33	7.2	110000	6 ABO69245_18	Continuation (19 o
C	41	33	7.2	319630	6 ABO67194	ABO67194 Listeria
C	42	32.6	7.1	351	7 ABX54413	ABX54413 Bovine ES
C	43	32.6	7.1	967	5 AAS74657	AAS74657 DNA encod
C	44	32.6	7.1	1476	6 AAS97253	AAS97253 Neisseria
C	45	32.6	7.1	1476	6 AAS97288	AAS97288 Neisseria

ALIGNMENTS

RESULT 1
AAC64620
ID AAC64620 standard; DNA; 456 BP.
XX
AC AAC64620;
XX
DT 26-FEB-2001 (first entry)
XX
DE Escherichia coli CsgB DNA sequence SEQ ID NO:4.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Escherichia coli.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UTVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36344.
XX
DR Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 134; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgB and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Sequence 456 BP; 146 A; 78 C; 113 G; 119 T; 0 U; 0 Other;
 Query Match
 Best Local Similarity 100.0%; Score 456; DB 3; Length 456;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAACC 60
 DB 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAACC 60
 QY 61 GCAGCAGGTTATGATTTAGTATTTAAGATTTAATTTCTGGCGCTGGGATTGCAACC 60
 DB 61 GCAGCAGGTTATGATTTAGTATTTAAGATTTAATTTCTGGCGCTGGGATTGCAACC 60
 QY 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120
 DB 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 120
 QY 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
 DB 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
 QY 181 CGCAGCAGGAGGCTCAAACTTTGGCGCTGGTGGCAGAAAGTATGACAGCGGCA 240
 DB 181 CGCAGCAGGAGGCTCAAACTTTGGCGCTGGTGGCAGAAAGTATGACAGCGGCA 240
 QY 241 AAGATTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 300
 DB 241 AAGATTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 300
 QY 301 GATCCAGATTTTCGCAAGGCTTATGATTTAATTTCTGGCGCTGGGATTGCAACC 300
 DB 301 GATCCAGATTTTCGCAAGGCTTATGATTTAATTTCTGGCGCTGGGATTGCAACC 300
 QY 361 GGTAAATTAAGCAATTTTACAGATGATGATCTCAAAAACGCAATTGATGACAGAG 420
 DB 361 GGTAAATTAAGCAATTTTACAGATGATGATCTCAAAAACGCAATTGATGACAGAG 420
 QY 421 CAGTCCCAATGCTATTTCGCTGACACAGCTTAA 456
 DB 421 CAGTCCCAATGCTATTTCGCTGACACAGCTTAA 456

RESULT 2
 AAC64618
 ID AAC64618 standard; DNA; 456 BP.
 AC AAC64618;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE *Salmonella enteritidis* AgfB DNA sequence SEQ ID NO:2.
 XX
 KW *Salmonella*; *agfA*; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 XX
 PN W0200060102-A2.

PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000MO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX (UWI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay MW,
 DR WPI; 2000-672631/65.
 DR P-PSDB; AAB36342.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PS protein useful for eliciting immune response in animal.
 PS Disclosure; Page 134; 139p; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TFP) nucleation depended
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Cgfa and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombinant Agfa gene into the chromosome of the
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Sequence 456 BP; 149 A; 87 C; 109 G; 111 T; 0 U; 0 Other;
 Query Match
 Best Local Similarity 73.0%; Score 332.8; DB 3; Length 456;
 Matches 379; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

QY 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAACC 60
 DB 1 ATGAAAAACAATTTGTTATTTATGATTTAACAATTTCTGGCGCTGGGATTGCAACC 60
 QY 61 GCAGCAGGTTATGATTTAGTATTTAAGATTTAATTTCTGGCGCTGGGATTGCAACC 120
 DB 61 GCAGCAGGTTATGATTTAGTATTTAAGATTTAATTTCTGGCGCTGGGATTGCAACC 120
 QY 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
 DB 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 180
 QY 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
 DB 121 TCTTCAATTTAATCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 240
 QY 181 CGCAGCAGGAGGCTCAAACTTTGGCGCTGGTGGCAGAAAGTATGACAGCGGCA 300
 DB 181 CGCAGCAGGAGGCTCAAACTTTGGCGCTGGTGGCAGAAAGTATGACAGCGGCA 300
 QY 241 AAGATTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420
 DB 241 AAGATTGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 420
 QY 301 GATCCAGATTTTCGCAAGGCTTATGATTTAATTTCTGGCGCTGGGATTGCAACC 360
 DB 301 GATCCAGATTTTCGCAAGGCTTATGATTTAATTTCTGGCGCTGGGATTGCAACC 360

```
Db 301 GATGCAATATATGCAAGGCGCTTACGCTATATAGTCAGCATTTATATCAGAAAGCTTCT 360
Oy 361 GGTATATAAGCAATATTATACACAGTATGTCTCAAAAACGCGCAATTGTAGTCAGAGA 420
Db 361 GGAATATAAGGCAATATTATACCGTACAGTACCGTACGCAAAAACAGAGTGTAGTCAGAAA 420
Oy 421 CAGTCGCAATATGCGGTATGCGGTGACACAGCTTAA 456
Db 421 CAGTCGCAATATGCGGTATGCGGTGACACAGCTTAA 456

RESULT 3
ACD68810 ID ACD68810 standard; DNA; 100 BP.
AC  ACD68810;
XX
AC  ACD68810;
XX
DT  18-SEP-2003 (first entry)
XX
DE  E. coli K12 MG1655 biochip probe SEQ ID 80.
XX
KM  Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS  Escherichia coli.
XX
PN  EPI260592-A1.
XX
PD  27-NOV-2002.
XX
PF  17-MAY-2001; 2001EP-00112179.
XX
PR  17-MAY-2001; 2001EP-00112179.
XX
PA  (MMGB-) MWG-BIOTECH AG.
XX
PI  Donner H, Drescher B, Huber A, Weber J;
XX
DR  WPI; 2003-241155/24.
XX
PT  Biochip containing probes complementary with open reading frames in
PT  Escherichia coli K12, useful for detecting gene expression and expression
PT  patterns.
XX
PS  Claim 3; Page 23; 2004pp; German.
XX
CC  This invention describes a novel biochip comprising probe spots, each
CC  containing many identical probes. The probes are nucleotide sequences of
CC  30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC  least one includes a segment of at least 20 bases identical with, or
CC  complementary to, a segment of an open reading frame (orf) of Escherichia
CC  coli K12. The biochip is used for specific detection of gene expression
CC  in K12 and for determining the gene expression pattern, e.g. for
CC  diagnostic determination of which E. coli strains are present in the gut,
CC  and to determine the effects of e.g. growth media on gene expression. The
CC  biochip provides as comprehensive as possible detection of the K12
CC  genome, with simultaneous analysis of many different genes with a single
CC  device, and comparison of gene expression between K12 and its mutants or
CC  other E. coli strains in a single experiment. Apart from qualitative and
CC  quantitative information about gene expression, it also allows
CC  measurements of population densities for the various strains. The use of
CC  synthetic oligonucleotides for preparation of probes allows free
CC  variation in probe length and ensures high purity (and thus selectivity,
CC  reactivity and reproducibility); also synthetic probes are generally
CC  shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC  ACD81540 represent oligonucleotide probes used with the biochip described
CC  in the invention
XX
SQ  Sequence 100 BP; 27 A; 19 C; 26 G; 28 T; 0 U; 0 Other;
Query Match 21.9%; Score 100; DB 7; Length 100;
Best Local Similarity 100.0%; Pred. No. 9.5e-21;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy 266 ACCTTGATATATGATCAGCGCGGCACTGCGCAAGATGCCAGTATTTCCGAAGTCTT 325
Db 1 ACCCTTGATATATGATCAGCGCGGCGAGTGCACAGATGCGAGATTTTCGAAGTCTT 60
Oy 326 ATGGTATATACGCGATATATTCAGAAAGTTCTGTTAA 365
Db 61 ATGGTATATACGCGATATATTCAGAAAGTTCTGTTAA 100

RESULT 4
ACD68812 ID ACD68812 standard; DNA; 100 BP.
AC  ACD68812;
XX
AC  ACD68812;
XX
DT  18-SEP-2003 (first entry)
XX
DE  E. coli K12 MG1655 biochip probe SEQ ID 82.
XX
KM  Biochip; gene expression; gut; diagnostic; detection; probe; ss.
XX
OS  Escherichia coli.
XX
PN  EPI260592-A1.
XX
PD  27-NOV-2002.
XX
PF  17-MAY-2001; 2001EP-00112179.
XX
PR  17-MAY-2001; 2001EP-00112179.
XX
PA  (MMGB-) MWG-BIOTECH AG.
XX
PI  Donner H, Drescher B, Huber A, Weber J;
XX
DR  WPI; 2003-241155/24.
XX
PT  Biochip containing probes complementary with open reading frames in
PT  Escherichia coli K12, useful for detecting gene expression and expression
PT  patterns.
XX
PS  Claim 3; Page 23; 2004pp; German.
XX
CC  This invention describes a novel biochip comprising probe spots, each
CC  containing many identical probes. The probes are nucleotide sequences of
CC  30-80 bases, are prepared ex situ from synthetic oligonucleotides and at
CC  least one includes a segment of at least 20 bases identical with, or
CC  complementary to, a segment of an open reading frame (orf) of Escherichia
CC  coli K12. The biochip is used for specific detection of gene expression
CC  in K12 and for determining the gene expression pattern, e.g. for
CC  diagnostic determination of which E. coli strains are present in the gut,
CC  and to determine the effects of e.g. growth media on gene expression. The
CC  biochip provides as comprehensive as possible detection of the K12
CC  genome, with simultaneous analysis of many different genes with a single
CC  device, and comparison of gene expression between K12 and its mutants or
CC  other E. coli strains in a single experiment. Apart from qualitative and
CC  quantitative information about gene expression, it also allows
CC  measurements of population densities for the various strains. The use of
CC  synthetic oligonucleotides for preparation of probes allows free
CC  variation in probe length and ensures high purity (and thus selectivity,
CC  reactivity and reproducibility); also synthetic probes are generally
CC  shorter than probes prepared by polymerase chain reaction. ACD68731 to
CC  ACD81540 represent oligonucleotide probes used with the biochip described
CC  in the invention
XX
SQ  Sequence 100 BP; 28 A; 18 C; 27 G; 27 T; 0 U; 0 Other;
Query Match 21.9%; Score 100; DB 7; Length 100;
Best Local Similarity 100.0%; Pred. No. 9.5e-21;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 123 TTCAATTATACAGCAGCATATATGTCAAGTCGACATTAATATATAGTCTCAGTTACG 182
```

Db 1 TTTATTTATATAGGACCAATATGTGCAAGTGGAGCTATATAGTGTCTGATTACG 60
 QY 183 GCAGGAGGCTCAAACTTTGGCGGTGTGTGGCAAGAA 222
 Db 61 GCAGGAGGCTCAAACTTTGGCGGTGTGTGGCAAGAA 100

RESULT 5

ACD68811
 ID ACD68811 standard; DNA; 100 BP.
 AC ACD68811;
 XX
 DT 18-SEP-2003 (first entry)
 XX
 DE E. coli K12 MG1655 biochip probe SEQ ID 81.
 XX
 KM Biochip; gene expression; gut; diagnostic; detection; probe; ss.
 XX
 OS Escherichia coli.
 XX
 PN EP1260592-A1.
 XX
 PD 27-NOV-2002.
 XX
 PF 17-MAY-2001; 2001EP-00112179.
 XX
 PR 17-MAY-2001; 2001EP-00112179.
 XX
 PA (MWGB-) MWG-BIOTECH AG.
 XX
 PI Donner H, Dreischer B, Huber A, Weber J;
 XX
 DR WPI; 2003-241155/24.
 XX
 PT Biochip containing probes complementary with open reading frames in
 PT Escherichia coli K12, useful for detecting gene expression and expression
 PT patterns.
 XX
 PS Claim 3; Page 23; 2004PP; German.

CC This invention describes a novel biochip comprising probe spots, each
 CC containing many identical probes. The probes are nucleotide sequences of
 CC 30-80 bases, are prepared ex situ from synthetic oligonucleotides and
 CC least one includes a segment of at least 20 bases identical with or
 CC complementary to, a segment of an open reading frame (orf) of Escherichia
 CC coli K12. The biochip is used for specific detection of gene expression
 CC in K12 and for determining the gene expression pattern, e.g. for
 CC diagnostic determination of which E. coli strains are present in the gut,
 CC and to determine the effects of e.g. growth media on gene expression. The
 CC biochip provides as comprehensive analysis as possible detection of the K12
 CC genome, with simultaneous analysis of many different genes with a single
 CC device, and comparison of gene expression between K12 and its mutants or
 CC other E. coli strains in a single experiment. Apart from qualitative and
 CC quantitative information about gene expression, it also allows
 CC measurements of population densities for the various strains. The use of
 CC synthetic oligonucleotides for preparation of probes allows free
 CC variation in probe length and ensures high purity (and thus selectivity,
 CC reactivity and reproducibility); also synthetic probes are generally
 CC shorter than probes prepared by polymerase chain reaction. ACD68731 to
 CC ACD61540 represent oligonucleotide probes used with the biochip described
 CC in the invention
 XX
 SQ Sequence 100 BP; 31 A; 20 C; 29 G; 20 T; 0 U; 0 Other;

Query Match 21.9%; Score 100; DB 7; Length 100;
 Best Local Similarity 100.0%; Pred. No. 9.5e-21;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GTTGTGCGCGAAGAGTAGTACACCGGCAAGATTGACGACAGAGATTATAC 267
 Db 1 GTTGTGCGCGAAGAGTAGTACACCGGCAAGATTGACGACAGAGATTATAC 60

QY 268 CTTGCATATATTGATCAGCGCGGCAAGTGCACCAAGCAGTCCCA 307
 Db 61 CTTGCATATATTGATCAGCGCGGCAAGTGCACCAAGCAGTCCCA 100

RESULT 6

AAN40272
 ID AAN40272 standard; cDNA; 779 BP.
 XX
 AC AAN40272;
 XX
 DT 09-JAN-1992 (first entry)
 XX
 DE Sequence of a clone numbered as Ag16 which corresponds to the S antigen
 DE of isolate FC27.
 XX
 KM Malaria; vaccine; antigen; epitope; immune response; ss.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 FH CDS 1..777
 FT 1..777
 FT misc_feature 1..8
 FT 1..8
 FT repeat_region 10..768
 FT 10..768
 FT /tag= c
 FT /rpt unit= 10..42
 FT /note= "final base of certain codons may differ but AA SQ
 FT 772..779
 FT encoded stays the same"
 FT /tag= b
 FT /label= linker
 PN WO8402917-A.
 XX
 PD 02-AUG-1984.
 XX
 PF 27-JAN-1984; 84WO-AU000016.
 XX
 PR 28-JAN-1983; 83AU-00007843.
 PR 10-JUN-1983; 83AU-00009788.
 PR 26-JAN-1984; 84AU-00023842.
 XX
 PA (HALT-) HALT INST MED RES.
 PA (KEMP/) KEMP D J.
 PA (AUME-) AUST INST MED RES.
 PA (SARA-) SARAMANE PTY LTD.
 XX
 PI Kemp DJ, Anders RF, Coppel RL, Brown G, Saint R, Cowman AF;
 PI Mitchell GF;
 DR WPI; 1984-201410/32.
 DR P-PSDB; AAP40376.
 XX
 PT Expression of plasmodium falciparum poly(peptide(s) from CDNA - for use
 PT in immunisation against malaria infection.
 XX
 PS Claim 5; Fig 1; 62pp; English.

CC The inventors claim a DNA SQ which comprises a nucleotide sequence
 CC corresp. to all or part of Plasmodium falciparum RNA. Also claimed is a
 CC synthetic peptide or polypeptide displaying the antigenicity of all or
 CC part of a P. falciparum antigen. The base sequence of clone Ag16 (see
 CC AAN40272) indicates that the antigen coded for by this partial
 CC polynucleotide sequence has a homologous repeat structure of 11 AAs
 CC tandemly repeated 23 times
 XX
 SQ Sequence 779 BP; 294 A; 125 C; 217 G; 143 T; 0 U; 0 Other;

Query Match 10.0%; Score 45.8; DB 1; Length 779;

Best Local Similarity 47.7%; Pred. No. 0.0011;
Matches 134; Conservative 0; Mismatches 147; Indels 0; Gaps 0;
Qy 126 ATTTAATCAGCAGCATTAATGCTCAAGCTGAGCTAATTAATAGTCTCAGTACGCCA 185
Db 198 ATTGAAGATCTCTCAAAAGCTAGTCAAGGTGATTAGAAATCTCTGCAAAAGCCAAATCA 257
Qy 186 GGGAGGCTCAAACTTTTGGCGGTGTTGGCGCAAGAGTAGTAGCAACCGGCAAAAT 245
Db 258 AGGAGGATTAGAAATCTCTGCAAAAGCTAGTCAAGGTGATTAGAAATCTCTGCAAAAGC 317
Qy 246 TGACCAAGCAGAGATTATTAACCTTGATATATGATCAGCGGCGAGTGCACAGATGC 305
Db 318 TAGTCAGAGAGATTAGAAATCTCTGCAAAAGCTAGTCAAGGTGATTAGAAATCTCTGC 377
Qy 306 CAGTATTTCCGAGGTCTTAATGTAATATCTGCGATGATTATCCAGAAAGTTCTGCTAA 365
Db 378 AAAAGCTAGTCAAGGTGATTAGAAATCTCTGCAAAAGCTAGTCAAGGTGATTAGAAAG 437
Qy 366 TAAAGCAAAATTTACACAGTATGCTCAAAAGCGCAA 406
Db 438 TCTGCAAAAGCTAGTCAAGGTGATTAGAAATCTCTGCAA 478

RESULT 7

ADA1934
ID ADA1934 standard; DNA; 4860 BP.

AC ADA1934;

DT 20-NOV-2003 (first entry)

DE DNA encoding Acinetobacter baumannii protein #3221.

KM de; gene; Acinetobacter baumannii; bacterial disease; antibacterial;
KW vaccine; plant biocontrol agent.

OS Acinetobacter baumannii.

PN US6562958-B1.

PD 13-MAY-2003.

PF 04-JUN-1999; 99US-00328352.

PR 09-JUN-1998; 98US-0088701P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Breton G, Bush D;

DR WPI; 2003-576092/54.

DR P-PSDB; ADA36060.

PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents
for diagnosing a bacterial disease, as components of antibacterial
PT vaccines, as targets for antibacterial drugs, or as biocontrol agents for
PT plants.

PS Example; SEQ ID NO 3221; 328BP; English.

XX The invention relates to isolated Acinetobacter baumannii nucleic acids.

XX CC The A baumannii nucleic acids and polypeptides are useful as reagents

XX CC for diagnosing a bacterial disease, as components of antibacterial

XX CC vaccines, as targets for antibacterial drugs, to detect the presence of

XX CC A. baumannii and other Acinetobacter species in a sample, in screening

XX CC compounds for the ability to interfere with the A. baumannii life cycle

XX CC or to inhibit A. baumannii infection, and as biocontrol agents for

XX CC plants. The present sequence represents DNA encoding an A. baumannii

XX CC protein.

XX SQ Sequence 4860 BP; 1639 A; 752 C; 998 G; 1471 T; 0 U; 0 Other;

Query Match 8.4%; Score 38.2; DB 8; Length 4860;
Best Local Similarity 45.9%; Pred. No. 0.51; Length 4860;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;
Qy 106 AATGAATGAGTAAGTCTTCAATTAATCAGCAGCAGCATTAATGCTCAAGCTGAGCTAAT 165
Db 3445 AATACTATGTTAATTAATTAATTAATGATGATGATGATGATGATGATGATGATGATG 3504
Qy 166 AATAGTCTCAGTACGCGAGGAGGCTCAAAACTTTTGGCGGTGTTGGCGCAAGAGT 225
Db 3505 ATTCAACTATTTATGTTGGTGGAGGCAATGATCGATCAAGGTGTTGATCAGAAATGAT 3564
Qy 226 AGTAGCAACCGGCAAGATTGACAGCAGAGATTATTAACCTTGATATATGATGATGATG 285
Db 3665 TATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3624
Qy 286 GCGGCGAGTGCACAGATGCGAGTATTTCCGAGGTCTTAATGATGATGATGATGATGATG 345
Db 3625 TATGTTGTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3684
Qy 346 ATCCAGAAAGTTTGTGTTAATTAAGCAATTTTACACAGTATG 388
Db 3685 TTACAGGTCAAGCTGGGAATGACATTAATTAATTAAGTAAAG 3727

RESULT 8

ADA71938/c
ID ADA71938 standard; DNA; 2000 BP.

AC ADA71938;

DT 20-NOV-2003 (first entry)

DE Rice gene, SEQ ID 5263.

KM Plant; bacterial infection; fungal infection; viral infection; rice;

KW gene; ds.

OS Oryza sativa.

PN MO2003000898-A1.

PD 03-JAN-2003.

PF 22-JUN-2001; 2001WO-IB001105.

PR 22-JUN-2001; 2001WO-IB001105.

PA (SYGN) SYNGENTA PARTICIPATIONS AG.

PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,

PI Katsagiri F, Qian S, Tao Y, Whitham S, Xie Z, Zhu T, Zou G;

DR WPI; 2003-175290/17.

PT Identifying at least one gene involved in plant resistance or response to
PT pathogenic infection for conferring resistance or tolerance to a plant to
PT bacterial, fungal or viral infection by determining or detecting plant
PT gene expression.

PS Claim 27; SEQ ID NO 5263; 899BP; English.

XX The present invention relates to a method (M1) for identifying genes

XX CC involved in plant resistance or response to pathogenic infection. M1

XX CC comprises identifying a gene whose expression is significantly altered in

XX CC the incompatible interaction of plant gene expression relative to

XX CC expression of the gene in an uninfected plant, in a mutant plant that

XX CC does not express a gene associated with response to pathogenic infection,

XX CC or in a corresponding incompatible or compatible interaction. (M1) is

XX CC useful for conferring resistance to resistance or tolerance to a plant to

XX CC bacterial, fungal or viral infection. The present sequence was used to

XX CC illustrate the invention.

XX XX Disclosure; Page 138; 139pp; English.

PS CC The present invention describes a recombinant agfa gene (I) where a

XX CC segment of the gene has been replaced by a segment of a foreign DNA

CC CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fibrillae (SEB17/TAF) nucleation depended

CC assembly system of strains of *Salmonella*, *Escherichia coli* and

CC *Enterobacteriaceae* for the production of fibrillae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fibrin subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene

CC back into the chromosome of the homologous species, replacing the native

CC copy of that gene; and (4) eliciting an immune response in an animal,

CC comprising separating an amino acid polymer comprising a recombinant Agfa

CC protein containing a replacement segment or segments of foreign amino

CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or

CC *Enterobacteriaceae* host cell, from the host cell and introducing the

CC polymer into the animal in conjunction with a carrier or diluent. (I) is

CC useful for the expression of recombinant Agfa protein which is useful for

CC eliciting an immune response in an animal. In a fibrillar presentation

CC system the heterologous antigens are presented in high numbers (up to

CC 500,000 copies/cell), the hybrid fibrin protein possesses both the

CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fibrillar subunit proteins are usually strong

CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fibrillae are easy and

CC inexpensive to purify in large amount. The present sequence is given in

CC the exemplification of the present invention

XX SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 8.0%; Score 36.6; DB 3; Length 456;

Best Local Similarity 54.0%; Pred. No. 0.61;

Matches 75; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 275 ATATTGATCAGCGCGGCGGAGTCCAGATGATTTCCAGAGTCTTATGTAATA 334

DB 206 AATGCGCATGCGAGTATGTAACGGCCGATGTAGCCAGGTCGGATATAGTA 265

QY 335 CTGCGATGATTAATCCAGAAAGTTCTGTAAATTAAGCAATATTATACAGATAGTACTC 394

DB 266 CTATTGAAGTCACTCAGAAATGTTTCAGAAATTAATGACCATGACAGTGGAACGCTA 325

QY 395 AAAAAACGCGCAATTGACT 413

DB 326 AAAAATCGATATTACTGT 344

RESULT 11

AAK34648/C

ID AAK34648 standard; DNA; 4890 BP.

XX AC AAK34648;

XX DT 17-OCT-2003 (revised)

XX DT 05-JUL-1999 (first entry)

DE Promoter containing sequence of SBE I.

XX XX Starch biosynthetic pathway; cereal plant; enzyme; SBE; SSS; DBE; GBS;

KM starch branching enzyme; starch soluble synthase; debranching enzyme;

KM endosperm; wheat; barley; granule-bound synthase; glutenin; starch;

KM grain softness protein I; bacterial isoamylase; glycogen synthase;

XX wSBE I-D4 gene; promoter; ss.

XX OS *Aegilops tauschii*.

XX PN MO914314-A1.

XX PD 25-MAR-1999.

XX PF 11-SEP-1998; 98WO-AU000743.

XX XX 12-SEP-1997; 97AU-00009108.

PR 20-MAR-1998; 98AU-00002509.

XX XX (CSIR) COMMONWEALTH SCI & IND RES ORG.

PA (AUS) UNIV AUSTRALIAN NAT.

PA (GOOD-) GOODMAN FIELDER LTD.

PA (LIMA-) GRP LIMAGRAIN PACIFIC PTY LTD.

XX Li Z, Morell M, Rahman S;

XX WPI; 1999-229525/19.

PT New isolated cereal plant enzyme genes used for, e.g. expression of

PT antisense sequences of granule bound synthase.

PS Claim 17; Page 63-65; 171pp; English.

XX XX The invention relates to a novel enzyme of starch biosynthetic pathway in

CC a cereal plant, where the enzyme is selected from starch branching enzyme

CC (SBE) I, SBE II, starch soluble synthase (SSS) I, and debranching enzyme

CC (DBE), with the proviso that the enzyme is not SSS I of rice, or SBE I of

CC rice or maize. The methods and products can be used for targeting

CC expression specifically to the endosperm of the seeds of cereal plants

CC such as wheat or barley. They can be used for the expression of e.g.

CC antisense sequences of granule-bound synthase (GBS), SBE II, low mol.

CC wt. glutenin, grain softness protein I, bacterial isoamylase, bacterial

CC glycogen synthase, and wheat high mol. wt. glutenin Bx1. They can be

CC used for modifying the characteristics of starch produced by a plant. The

CC present sequence represents the wheat SBE I promoter containing sequence.

XX (Updated on 17-OCT-2003 to standardise OS field)

XX SQ Sequence 4890 BP; 1403 A; 1067 C; 1035 G; 1385 T; 0 U; 0 Other;

Query Match 7.7%; Score 35.2; DB 2; Length 4890;

Best Local Similarity 55.8%; Pred. No. 4.3;

Matches 87; Conservative 0; Mismatches 68; Indels 1; Gaps 1;

QY 258 AGATTATTAACCTTGATATATATGATCGGCGGAGTCCAGATGCGGATTTGGCA 317

DB 1097 AGAAGAAACCTTCCATAAATGAACAAACAGACGCGCAAGTATGCTTTGGCA 1038

QY 318 AGGTGCTTAATGTAATCTGCGA-TGATTAATCCAGAAAGTTCTGTAAATGAACAATA 376

DB 1037 ACTCATACATTTGAAATTTGTTAGTTAAATGAACTAAGTTCTTATTAGATGCA 978

QY 377 TTACACAGATGTAATCAAAAACGCGCAATTGTAG 412

DB 977 TTGCATTGCATTTACTAGTACCCGGAATTTGAG 942

RESULT 12

AAO62647

ID AAO62647 standard; DNA; 646 BP.

XX AC AAO62647;

XX DT 25-MAR-2003 (revised)

XX DT 18-OCT-1994 (first entry)

DE Fibronectin binding protein fnda gene.

XX XX Fibronectin binding protein; FBP; curli pill; pilus;

KM immunization; vaccine; fnda gene; ss.

XX OS *Escherichia coli*.

XX PN MO9410330-A1.

XX PD 11-MAY-1994.

XX PF 03-NOV-1993; 93WO-US010547.

PR 03-NOV-1992; 92US-00970846.
 XX (REAT/) REA T S.
 PA
 XX Normark S, Olsen A;
 PI
 XX WPI, 1994-167484/20.
 DR
 XX P-PSDB; AAR52664.
 PT New fibronectin binding protein curlin from *E. coli* - used for the
 PT treatment of infections caused by *E. coli* and for immunisation against *E.*
 PT *coli* infection.
 XX
 PS Disclosure; Fig 2; 33pp; English.
 CC
 CC The *fbaA* gene of *E. coli* cattle fecal isolate A012 includes DNA encoding
 CC a new fibronectin binding protein (fBP), curlin, which is produced as
 CC curli pili when expressed in *E. coli* (pFnb20) DSM 4585. The *fbaA* gene
 CC sequence from pFnb20 is given in AA062647, and includes the sequence
 CC encoding the 17 kDa curlin subunit (amino acids 21-142 of AAR52664).
 CC (Updated on 25-MAR-2003 to correct PN field.)
 CC
 XX
 SQ Sequence 646 BP; 168 A; 146 C; 161 G; 171 T; 0 U; 0 Other;
 Query Match 7.7%; Score 35; DB 2; Length 646;
 Best Local Similarity 100.0%; Pred. No. 2.2;
 Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 416 AGAGACAGTCGCAATGCGCTATTGGGTGACACAA 450
 Db 1 AGAGACAGTCGCAATGCGCTATTGGGTGACACAA 35
 RESULT 13
 ID AAQ73066 standard; DNA; 361 BP.
 XX
 AC AAQ73066;
 XX
 DT 27-AUG-2003 (revised)
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE Agfa sequence.
 XX
 KW Salmonella; Agfa; vaccine; genetic immunization; ds.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..359
 FT /tag= a
 FT /note= "Agfa"
 FT /tag= d
 FT /note= "TAF5 primer (pair with TAF6)"
 FT /tag= b
 FT /note= "TAF3 primer (pair with TAF4)"
 FT /tag= e
 FT /note= "TAF6 primer (pair with TAF5)"
 FT complement (292..402)
 FT /tag= c
 FT /note= "TAF4 primer (pair with TAF3)"
 XX
 PN W09425598-A2.
 XX
 PD 10-NOV-1994.
 XX
 PF 26-APR-1994; 94WO-IB000207.
 XX
 PR 26-APR-1993; 93US-00054452.

XX
 PA (UVT-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX
 XX Kay WW, Collinson SK, Cloughier SC, Doran JL;
 XX WPI, 1994-358275/44.
 DR
 XX P-PSDB; AAR62761.
 PT Eliciting an immune response to Salmonella - using attenuated Salmonella
 PT strains, vector constructs, or compens. contg. fimbrial type proteins.
 XX
 PS Disclosure; Fig 7A; 95pp; English.
 CC
 CC The DNA encodes the Salmonella enteritidis27655-3b *trpH*A mutant strain
 CC *agfa* gene cloned into pUC19. The DNA and isolated proteins are used in
 CC genetic immunization and vaccine compositions, respectively, to elicit an
 CC immune response to Salmonella in animals (e.g. food producing animals)
 CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
 CC AUG-2003 to correct OS field.)
 CC
 XX
 SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
 Query Match 7.6%; Score 34.6; DB 2; Length 361;
 Best Local Similarity 53.3%; Pred. No. 2.3;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGCGGACGATGCGATGTTTGGCAAGTGTATGTAATACT 336
 Db 145 ATTACCCAGAGCGGTTATGTTAAGCGCCGATGAGGCCAGGGTGGCATTAATAGTACT 204
 QY 337 GCGATGATTTATCCAGAAAGTTCTGTATATTAAGCAATATTATACAGTATGTAATCA 396
 Db 205 ATTGAACGTACTCAGAAAGTTTTCAGAAATATATGATCCACATCGACGATGAGACGTTAA 264
 QY 397 AAAACGCAATTGTACT 413
 Db 265 AACTCCGATATTACTGT 281
 RESULT 14
 ID AAT74141
 XX
 AC AAT74141;
 XX
 DT 25-MAR-2003 (revised)
 DT 30-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b *trpH*A mutant *agfa* gene fragment.
 XX
 KW Enteropathogenic bacteria; enterobacteria; *S. enteritidis*; antibody; ds.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..360
 FT /tag= a
 FT /label= *agfa*_gene_fragment
 FT /tag= b
 FT /label= Primer_TAF5
 FT /tag= c
 FT /label= Primer_TAF3
 FT /tag= d
 FT /label= Primer_TAF6
 FT complement (103..128)
 FT /tag= e
 FT /label= Primer_TAF4
 XX
 PN US5635617-A.

XX 03-JUN-1997.
 PD 26-APR-1994; 94US-00233788.
 XX 26-APR-1993; 93US-00054452.
 XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX Collinson SK, Kay WW, Doran JL;
 PI WPI; 1997-309886/28.
 DR P-PSDB; AAM23569.
 XX Isolated *Salmonella* gene *agfa* - used for diagnosis of *Salmonella* or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX Claim 1; Col 107-110; 85pp; English.
 XX The present sequence represents an isolated *agfa* gene fragment derived
 CC from *Salmonella enteritidis* 27655-3b Typhoid mutant strain. The nucleic
 CC acid can be used to provide diagnostic assays for *Salmonella* and/or
 CC enteropathogenic bacteria of the family Enterobacteria. It can also be
 CC used to provide proteins and antibodies which can be used for assays. The
 CC nucleic acid sequence can be used to provide probes or primers which can
 CC specifically hybridise to nucleic acid molecules from greater than 99% of
 CC *Salmonella* strains that are pathogenic to warm-blooded animals relative
 CC to nucleic acid molecules from virtually all other microbial organisms.
 CC (Updated on 25-MAR-2003 to correct PF field.)
 XX
 SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;

Query Match 7.6%; Score 34.6; DB 2; Length 361;
 Best Local Similarity 53.3%; Pred. No. 2.3;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGGCGAGTCCAGATGCCAGTATTTCGCAAGGCTTATGTAATCT 336
 DB 145 ATTACCCAGAGCGGTTATGTTAGTACGCCGCCATGTAGCCGAGGTGCGATTAATGTA 204
 QY 337 GCGATGATTATTCAGAAAGTTCTGTGTAATAAAGCAATATTATACACAGTATGTA 396
 DB 205 ATGAACTGACTCAGAAATGTTTCAGAAATATATGCAATGCACAGTGGAGCGTAA 264
 QY 397 AAAACGCAATTGTA 413
 DB 265 AACTCCGATATTACTGT 281

RESULT 15
 ID AA087467 standard; DNA; 456 BP.
 XX
 AC AA087467;
 XX
 DT 25-MAR-2003 (revised)
 DT 26-JUN-1995 (first entry)
 XX
 DE *Agfa* sequence.
 XX
 KM *Salmonella*; *Agfa*; vaccine; genetic immunization; ds.
 XX
 OS *Salmonella*.
 XX
 FH Key
 FH CDS 1.454
 FT Location/Qualifiers
 FT /tag= a
 FT /note= "Agfa"
 XX
 PN W09425598-A2.
 XX
 PD 10-NOV-1994.
 XX

PF 26-APR-1994; 94WO-IB000207.
 XX 26-APR-1993; 93US-00054452.
 XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PA (KING/) KING J.
 XX Kay WW, Collinson SK, Clouchier SC, Doran JL;
 PI WPI; 1994-358275/44.
 DR P-PSDB; AAR74625.
 XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compans. contg. fibrial type proteins.
 XX Disclosure; Fig 7B; 95pp; English.
 XX The DNA encodes the *Salmonella Agfa* protein. The DNA and isolated
 CC proteins are used in genetic immunization and vaccine compositions,
 CC respectively, to elicit an immune response to *Salmonella* in animals (e.g.
 CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX
 SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 7.6%; Score 34.6; DB 2; Length 456;
 Best Local Similarity 53.3%; Pred. No. 2.5;
 Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 277 ATTGATCAGCGGCGAGTCCAGATGCCAGTATTTCGCAAGGCTTATGTAATCT 336
 DB 208 ATTACCCAGAGCGGTTATGTTAGTACGCCGCCATGTAGCCGAGGTGCGATTAATGTA 267
 QY 337 GCGATGATTATTCAGAAAGTTCTGTGTAATAAAGCAATATTATACACAGTATGTA 396
 DB 268 ATGAACTGACTCAGAAATGTTTCAGAAATATATGCAATGCACAGTGGAGCGTAA 327
 QY 397 AAAACGCAATTGTA 413
 DB 328 AACTCCGATATTACTGT 344

Search completed: March 15, 2004, 17:51:50
 Job time : 248.684 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaatacaatcgtatc.....ctcgctgacacacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6C.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
C 1	39.6	8.7 7218 1	US-08-232-463-14	Sequence 14, Appl
2	38.2	8.4 4860 4	US-09-328-352-3221	Sequence 3221, Ap
3	34.6	7.6 361 1	US-08-233-788A-56	Sequence 56, Appl
4	34.6	7.6 456 1	US-08-233-788A-58	Sequence 58, Appl
5	31.8	7.0 580073 4	US-08-545-528D-1	Sequence 1, Appl1
6	31.8	7.0 640681 4	US-09-790-988-1	Sequence 1, Appl1
7	31.4	6.9 525 4	US-09-107-532A-2325	Sequence 2325, Ap
8	31.4	6.9 2785 4	US-08-956-171E-133	Sequence 133, App
9	31.4	6.9 53332 4	US-09-801-861-3	Sequence 32, Appl
10	31.4	6.8 62909 4	US-09-586-002-32	Sequence 32, Appl
11	31.2	6.8 606 4	US-09-540-236-1518	Sequence 1518, Ap
12	31.2	6.8 1404 4	US-09-134-001C-398	Sequence 398, App
13	31	6.8 2139 4	US-09-058-584-50	Sequence 50, Appl
14	31	6.8 1664976 4	US-08-916-421B-1	Sequence 1, Appl1
15	30.6	6.7 832 4	US-09-621-976-3813	Sequence 2813, Ap
16	30.6	6.7 1664976 4	US-08-916-421B-1	Sequence 1, Appl1
17	30.4	6.7 831 4	US-09-328-352-3901	Sequence 3901, Ap
18	30.4	6.7 1023 4	US-09-540-236-1556	Sequence 1556, Ap
19	30.4	6.7 1227 4	US-09-107-532A-2544	Sequence 2544, Ap
20	30.4	6.7 1622 4	US-08-956-171E-732	Sequence 732, App
21	30.4	6.7 1776 4	US-09-134-000C-3324	Sequence 3324, Ap
22	30.4	6.7 1879 4	US-09-023-655-661	Sequence 661, App
23	30.4	6.7 2001 3	US-08-714-918-82	Sequence 92, Appl
24	30.4	6.7 2001 3	US-09-265-315-92	Sequence 92, Appl
25	30.4	6.7 2001 3	US-09-265-315-92	Sequence 92, Appl
26	30.4	6.7 2001 3	US-09-266-417-92	Sequence 92, Appl
27	30.4	6.7 2001 4	US-09-528-709-92	Sequence 92, Appl

28	30.4	6.7 2001 4	US-09-527-745-92	Sequence 92, Appl
29	30.4	6.7 99629 4	US-09-586-002-37	Sequence 37, Appl
C 30	30.2	6.6 417 4	US-09-621-976-10111	Sequence 10111, A
31	30.2	6.6 2364 3	US-09-172-045-1	Sequence 1, Appl1
32	30.2	6.6 2364 4	US-09-342-325C-1	Sequence 1, Appl1
33	30.2	6.6 4215 4	US-09-620-312D-295	Sequence 295, App
34	30.2	6.6 13508 4	US-08-956-171E-120	Sequence 120, App
35	30.2	6.6 152321 3	US-09-128-155-16	Sequence 16, Appl
36	30.2	6.6 176373 3	US-09-128-155-17	Sequence 17, Appl
37	30	6.6 2136 2	US-08-949-941B-1	Sequence 1, Appl1
C 38	29.8	6.5 7091 3	US-08-975-762-46	Sequence 46, Appl
C 39	29.8	6.5 7091 3	US-09-295-028-46	Sequence 46, Appl
C 40	29.8	6.5 7091 4	US-09-106-582-46	Sequence 46, Appl
C 41	29.8	6.5 7091 4	US-09-159-469-46	Sequence 46, Appl
C 42	29.8	6.5 7091 4	US-09-693-542-46	Sequence 46, Appl
C 43	29.8	6.5 13158 2	US-08-687-080-105	Sequence 105, App
C 44	29.4	6.4 703 4	US-09-016-434-120	Sequence 120, App
C 45	29.2	6.4 369 4	US-09-328-352-3808	Sequence 3808, Ap

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHRIFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZ9pC-Fls
US-08-232-463-14
Query Match 8.7%; Score 39.6; DB 1; Length 7218;

Best Local Similarity 10.5%; Pred. No. 0.0088;

Matches 39; Conservative 166; Mismatches 165; Indels 0; Gaps 0;

```
QY 64 GCAGTTATGATTAAGTAACTTCAAGATATTAATCCGGTAATGAATGATAGTCT 123
DB 1523 GCATCTATTCAGTTTCAAAAACGCGATGTCATCTGATTAATCTATCTATGCA 1464
QY 124 TCATTAATCAAGGCGCATTAATGTCAGTGGACTGATTAATGTCCTAGTACG 183
DB 1463 GTAGTTAAAGATGAAAGATTTGTACRRRRRRRRRRRRRRRRRRRRRRRRRR 1404
QY 184 CAGGAGGCTCAAACTTTGGCGTGTGGCAGAGAGTACTGACACGGCGCAAG 243
DB 1403 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1344
QY 244 ATTACCAAGACAGAGATTAATCTGATTAATGATCAAGCGCGAGTGCAGCAT 303
DB 1343 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1284
QY 304 GCAGTATTTGCAAGTCTTATGTAATATCTGGATGATTCAGAAAGTTCTGT 363
DB 1283 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1224
QY 364 AATAAGCAAAATTCACAGATGCTCTCAAAAACGCAATTTGATGACAGACG 423
DB 1223 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1164
QY 424 TCGCAATG 433
DB 1163 RRRRRRRRR 1154
```

RESULT 2

```
US-09-328-352-3221
; Sequence 3221, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 3221
; LENGTH: 4860
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-3221
```

Query Match 8.4%; Score 38.2; DB 4; Length 4860;

Best Local Similarity 45.9%; Pred. No. 0.021;

Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```
QY 106 AATGAATGAGTAACTCTTCAATTAACAGCAGCATATATGTCAGTGGACTAT 165
DB 3445 AATCTATTTGTAATTAATTAAGTACGATGACGAGATATCTTCAAGCAGATGCTGAG 3504
QY 166 AATAGTCTAGTTACGCGAGGAGGCTCAAACTTTGGCGTGTGGCAGAAAGT 225
DB 3505 ATTCACTATTATGTTGGAGGCAATGATGATTCAGAGTGGTGTGACAAATGAT 3564
QY 226 AATGACAGGCGGCAAGATTAACAGACAGAGATTAATCTTGAATATGATGATG 285
DB 3565 TATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3624
QY 286 GCGGCGAGTGCACCAATGATTTGCGAAGTGTATGTAATGATGATGATGATG 345
DB 3625 TATGTTGTTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 3684
QY 346 ATCCAGAAAGTTCTGATTAATTAAGCAATATTAACAGATGATG 388
DB 3685 TTACAGAGTCAAGCTGGAGATGACACTATATATATATGATTAAG 3727
```

RESULT 3

```
US-08-233-788A-56
; Sequence 56, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, Karen S.
; APPLICANT: Cloutier, Sharon C.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
; NUMBER OF SEQUENCES: 61
; CORRESPONDENCE ADDRESS:
; ADDRESS: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C2
; TELEPHONE: (206) 682-4900
; TELEFAX: (206) 682-6031
; INFO: 3723836 SEDANBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 361 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..357
US-08-233-788A-56
```

Query Match 7.6%; Score 34.6; DB 1; Length 361;

Best Local Similarity 53.3%; Pred. No. 0.087;

Matches 73; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
QY 277 ATTGATCAAGGCGGACAGTGCACAGATGATTTCCAAAGTGTGATGATGAT 336
DB 145 ATTACCAAGAGCGGTATGTTAAGCGCGCGATGTCAGAGGTCGATTAATGAT 204
QY 337 GCGATGATTAACCAAGAGTGTGTTAATTAAGCAAAATATTAACAGATGATG 396
DB 205 ATTGAATGCTAGATGTTTCAAGAAATATGTCACATGACACAGTGAACGCTAAA 264
QY 397 AAAAGCGCAATGTTAGT 413
DB 265 AACTCGATATTAATCTGT 281
```

RESULT 4

```
US-08-233-788A-58
; Sequence 58, Application US/08233788A
; Patent No. 5635617
; GENERAL INFORMATION:
; APPLICANT: Doran, James L.
```

```

RESULT 5
US-08-545-528D-1
; Sequence 1, Application US/08545528D
; Patent No. 6537773
; GENERAL INFORMATION:
; APPLICANT: Fraser et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Mycoplasma Genitalium Genome, Fragment
; Patent No. 6537773
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: PB193p1
; CURRENT APPLICATION NUMBER: US/08/545,528D
; CURRENT FILING DATE: 1995-10-19
; PRIOR APPLICATION NUMBER: US 08/488,018
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/473,545

```

Query Match	7.0%;	Score 31.8;	DB 4;	length 580073;
Best Local Similarity	43.1%;	Pred. No. 30;		
Matches 153;	Conservative	0;	Mismatches 202;	Indels 0;
			Gaps	0;

```

RESULT 6
US-09-790-988-1
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: MATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHITA
; APPLICANT: SAKAKI, YOSHITYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APRIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: J22000-107160
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

```

	Query Match	Similarity	Score	DB	Length
Best Local	99%	46.9%	31.8	DB 4	640661
Matches	99	Conservative	0	Mismatches	112
				Indels	0
				Gaps	0

	Query	DB	Score	Length
Qy	2 TGAAGAAACAAATGTTATTTATGATGTTAACTACTGGGTGCGCTGGGATTCAGCCG	61		
Db	79853 TTGACAAAGAGACGTGAGATTAATTTAAACAAAAAGCATRTGAAACAGGTTTCC	79912		
Qy	62 CACGAGGTTATGATTAGCTAATTCAGATATTAATCTCGCGGTAAATGTAATGAGTAAGT	121		
Db	79913 AAAACAGACTTCTTAAAAAGCCAAAGAGAACACTTTGTTGAAAAATTAATTAATGATT	79972		

QY 122 CTCATTATTCAGGAGCCATATGTCAGCTGGAGCTATATATGCTCAGTTAC 181
 Db 79973 TATTTTAAATTTTGAATAATTCCTGCTGAATTTGAATAATGACTATGTTTCAATTTAC 80032
 QY 182 GCGAGGAGAGCTCAAAATTTGGGGCTGT 212
 Db 80033 TAAAAACAGTTTAAATAATTCATTTATGT 80063

RESULT 7

US-09-107-532A-2325
 ; Sequence 2325 Application US/09107532A
 ; Patent No. 6583275
 ; GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
 NUMBER OF SEQUENCES: 7310
 CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION
 STREET: 100 Beaver Street
 CITY: Waltham
 STATE: Massachusetts
 COUNTRY: USA
 ZIP: 02354

COMPUTER READABLE FORM:
 MEDIUM TYPE: CD-ROM ISO9660
 COMPUTER: PC
 OPERATING SYSTEM: <Unknown>
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/107, 532A
 FILING DATE: 30-Jun-1998
 PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 60/085, 598
 FILING DATE: 14 May 1998
 APPLICATION NUMBER: 60/051571
 FILING DATE: July 2, 1997
 ATTORNEY/AGENT INFORMATION:

NAME: Arinello, Pamela Deneke
 REGISTRATION NUMBER: 40,489
 REFERENCE/DOCKET NUMBER: GTC-012
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (781)893-8277
 TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 2325:
 SEQUENCE CHARACTERISTICS:

LENGTH: 525 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: circular
 MOLECULE TYPE: DNA (genomic)
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium
 FEATURE:

NAME/KEY: misc.feature
 LOCATION: (B) LOCATION 1...525
 SEQUENCE DESCRIPTION: SEQ ID NO: 2325:
 US-09-107-532A-2325

Query Match
 Best Local Similarity 54.9%; Score 31.4; DB 4; Length 525;
 Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 270 TGCATATTTGATCGCGGAGTCCAGACGATGCGCATTTTTCGCAAGTCTTATG 329
 Db 213 TGAATATTCGATTAAGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 329
 QY 330 TAAATCTGCATGATATTCAGAAAGGTTCTGTAATTAAGCAAAATATTAC 382
 Db 273 CTATATTTGATGAGTGTGAATCAATTAATGCGATTAAGCAAAATATTGTC 325

RESULT 8
 US-08-956-171E-133/C
 ; Sequence 133 Application US/08956171E
 ; Patent No. 659114
 ; GENERAL INFORMATION:

APPLICANT: Charles Kunach
 Gail H. Choi
 Patrick S. Dillon
 Craig A. Rosen
 Steven C. Barash
 Michael R. Fannon
 TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 5256
 CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/956, 171E
 FILING DATE: 20-Oct-1997
 CLASSIFICATION: <Unknown>
 PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 60/009, 861
 FILING DATE: January 5, 1996
 APPLICATION NUMBER: 08/781, 986
 FILING DATE: January 3, 1997
 ATTORNEY/AGENT INFORMATION:

NAME: Mark J. Hyman
 REGISTRATION NUMBER: 46,789
 REFERENCE/DOCKET NUMBER: PB248P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (240) 314-1224
 TELEFAX: (301) 309-8439

INFORMATION FOR SEQ ID NO: 133:
 SEQUENCE CHARACTERISTICS:

LENGTH: 2785 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 133:
 US-08-956-171E-133

Query Match
 Best Local Similarity 49.7%; Score 31.4; DB 4; Length 2785;
 Matches 80; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 125 CATTATTCAGGAGCCATATGTCAGCTGGAGCTATATATGCTCAGTTAC 184
 Db 2531 CAATTAATCAAGAGATATCATTTTGAACCGGTTGCTTGTAGTGTTCAGGTACCA 2472
 QY 185 AGGAGGCTCAAACTTTGGCGGTTGTCGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 244
 Db 2471 ATATGTCACCACTTAATATTCGCTTAATGATTAACAACAGATAGGTGTTATACCAAAA 2412
 QY 245 TTACACAGACAGAGATTAATACCTTGATATATTTGATCG 285
 Db 2411 TTGCAACTGTATGCAAAATTTATCTGCCAAGTTAGCACAG 2371

RESULT 9
 US-09-801-861-3
 ; Sequence 3 Application US/09801861
 ; Patent No. 6492154

GENERAL INFORMATION:
APPLICANT: YAN, Chunhua et al.
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
TITLE OF INVENTION: THEROPOF
FILE REFERENCE: CL001098
CURRENT APPLICATION NUMBER: US/09/801,861
CURRENT FILING DATE: 2001-03-09
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO: 3
LENGTH: 53332
TYPE: DNA
ORGANISM: Human
US-09-801-861-3

Query Match 6.9%; Score 31.4; DB 4; Length 53332;
Best Local Similarity 61.7%; Pred. No. 12;
Matches 50; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 336 TGCATGTTATCCGAAAGTTCGTATTAACCAATTTACACGTATGTAATCA 395
Db 8632 TGTATAGATATCAACAAGATATGATATTAAGCACTTATCACAGAGTCA 8691
Qy 396 AAAACGCAATTGTAGTCA 416
Db 8692 GAGTTCAACCAATGTGACCA 8712

RESULT 10
US-09-596-002-32/c
Sequence 32, Application US/09596002
Patent No. 6632636
GENERAL INFORMATION:
APPLICANT: Lagace, Robert, E.
APPLICANT: Paterson, Chandra
APPLICANT: Berg, Kim, L.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
FILE REFERENCE: PM-0008-4 US
CURRENT APPLICATION NUMBER: US/09/596,002
CURRENT FILING DATE: 2000-06-16
PRIOR APPLICATION NUMBER: 60/140,121
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PERL Program
SEQ ID NO 32
LENGTH: 62909
TYPE: DNA
ORGANISM: M. catarrhalis
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: incyte template ID No. 6632636 32
PUBLICATION INFORMATION:
US-09-596-002-32

Query Match 6.9%; Score 31.4; DB 4; Length 62909;
Best Local Similarity 49.7%; Pred. No. 14;
Matches 80; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 221 AAGTAGTACGACCGGCAAGATTGACGACAGAGATTAATTAACCTTCATATATTG 280
Db 8029 AATTGATTAACACCATATGAGATATGTAAGAAAGTTATCTTAATTCATTTATGG 7970
Qy 281 ATCAGCGCGGAGTCCACAGATGCAATTTGGCAAGTGTATGTAATTAATCTGCGA 340
Db 7869 CGGTGGCGAGTTTGTCTAACCGAGACTTATGAGCTTATCATCAACCAATGGCG 7910
Qy 341 TGATTATCCAGAAAGTTCTGTATATAAGCAATATTACA 381
Db 7909 GTTTAGCATTTGACATTTTGGACCTTCACCAATGTGCGA 7869

RESULT 11

US-09-540-236-1518
Sequence 1518, Application US/09540236
Patent No. 6673910
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
FILE REFERENCE: 2709,2005-001
CURRENT APPLICATION NUMBER: US/09/540,236
CURRENT FILING DATE: 2000-04-04
NUMBER OF SEQ ID NOS: 3840
SEQ ID NO 1518
LENGTH: 606
TYPE: DNA
ORGANISM: M. catarrhalis
US-09-540-236-1518

Query Match 6.8%; Score 31.2; DB 4; Length 606;
Best Local Similarity 50.0%; Pred. No. 1.5;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 226 AGTAGCAACCGGCAAGATTGACGACAGAGATTATTAACCTTCATATATTGATCAG 285
Db 4 ATTAACAACCATATGAGATATGTAAGAAAGATTATCTTAATTTGATTTGCGCGTG 63
Qy 286 GCGGCGAGTGGCAACGATGCGAGTATTTGCAAGTGTATGTAATTAATCTGCGATATT 345
Db 64 GCAGGTTTGTCTAACGAGCGACTTATGAGCTTATCTTATATCAACCAATGCCGTTT 123
Qy 346 ATCCAGAAAGTTCTGTATTAATTAAGCAATATTACA 381
Db 124 AGCATTAACATTTTGGCAGCTCTCAACCAATGTGCA 159

RESULT 12
US-09-134-001C-398
Sequence 398, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucelte-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 398
LENGTH: 1404
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-398

Query Match 6.8%; Score 31.2; DB 4; Length 1404;
Best Local Similarity 52.3%; Pred. No. 2.3;
Matches 69; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 275 ATATTATCAAGGCGGAGTCCCAAGATGCGATTTTGGCAAGTGTATGTAATA 334
Db 1079 ATTCTGATTAAGAACGAGATGATGACTGATTAATAGTCTCTCAATCTTACCGCTACTT 1138
Qy 335 CTGCGATGATTTATCCAGAAAGTTCTGTATTAATTAAGCAATATTACACAGATGTAATC 394
Db 1139 CAATTAATTAATTAAGATTAATGATGACCAATATACTCAATTTATCAACCAATATCAC 1198
Qy 395 AAAAAGCGCA 406
Db 1199 AAGATTAATGCA 1210

RESULT 13
US-09-059-584-50
Sequence 50, Application US/09059584
Patent No. 6440701
GENERAL INFORMATION:
APPLICANT: Myers, Lisa E
APPLICANT: Schryvers, Anthony B
APPLICANT: Hartness, Robin E
APPLICANT: Loomore, Sheena M.
APPLICANT: Du, Run-Pan
APPLICANT: Yang, Yan-Ping
APPLICANT: Klein, Michel H
TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: 6th Floor, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/059,584
FILING DATE: 14-APR-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/778,570
FILING DATE: 03-JAN-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I
REGISTRATION NUMBER: 24973
REFERENCE/DOCKET NUMBER: 1038-794
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 2139 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-059-584-50

Query Match
Best Local Similarity 59.8%; Score 31; DB 4; Length 2139;
Matches 52; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 306 CAGTATTCGACAGTGTGTTATGTTATCTGCGATGATTTACGAGAAGTTCTGGTAA 365
DB 1845 CAGTATTCGACAGTGTGTTATGTTATCTGCGATGATTTACGAGAAGTTCTGGTAA 365
QY 366 TAAAGCAATATTTACACAGTATGTAC 392
DB 1905 TAAAGCAATATTTACACAGTATTTAAAC 1931

RESULT 14
US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanoco-
coccus jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B

CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g

```

LOCATION: (1349491)..(1349491)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
/ OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664855)
/ OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match 6.8%; Score 31; DB 4; Length 1664976;
Best Local Similarity 49.1%; Pred. No. 83;
Matches 82; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

239 CAAAGATTGACGACGACGAGATTATTAACCTTGACATATATGATGACGGCGGACGATGCCA 298
122859 CATATATATAGAGAGATGATGAACTACCGTGCTACTTATATACATGACGAGGAGACTG 1228000
299 ACGATGCCAGTATTTGCGAAGGTGCTTATGTGTAATACGATGATTCGAGAAAGTT 358
122799 GAGATTACATATATGCAATATATATGATGCTTACGCTGTTCAAAAATAGCATTAAT 1227400
359 CTGGTAATTAAGCAAAATATTACACGATGATGACTCAAAAAACGGCA 405
122739 ATACAAATGAGCATTTAACAGACAGTATGGGGTTTATAGAACTGCA 122693

RESULT 15
US-09-621-976-2813
/ Sequence 2813, Application US/09621976
/ Patent No. 6639063
/ GENERAL INFORMATION:
/ APPLICANT: Dumas Milne Edwards, J.B.
/ APPLICANT: Jobert, S.
/ APPLICANT: Giordano, J.Y.
/ TITLE OF INVENTION: ESTs and Encoded Human Proteins.
/ FILE REFERENCE: GENSET.054PR2
/ CURRENT APPLICATION NUMBER: US/09/621,976
/ CURRENT FILING DATE: 2000-07-21
/ NUMBER OF SEQ ID NOS: 19335
/ SOFTWARE: Patent.pm
/ SEQ ID NO 2813
/ LENGTH: 832
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 235..399
US-09-621-976-2813

6.7%; Score 30.6; DB 4; Length 832;
Best Local Similarity 16.5%; Pred. No. 2.8; Indels 195; Indels 2; Gaps 1;
Matches 69; Conservative 151; Mismatches 195; Indels 2; Gaps 1;

10 AAATGTATTATGATGTGTAACATACTGCGTGCGCTGGATTGACGCCGACGAGT 69
40 WRRKKKKAAWYKWTWTWYRVAMGWYKKKAMCRRTKKKKKKGYMMWMTWGMRSYV 99
70 TATGATTTAGCTAATTCGAATATPACTTCGGCGTAATGAATTGAGTAAGTCTTCA 129
100 AAMRTTWTGVAVYVSMWMTWYRVAMGWYKKKAAVYAKTTCYSKGTWTKRWKKAATWTKT 159

```

```

Qy 130 AATCAGGACCCATTAATGCTCAAGCTGGACTAATAATAGTCTCAGTTACGGCAGGA 189
Db 160 YWAATRYMMWMCWTKRMRASWNYCWMWGWKARKSTWRKRSYASARSASAKRCYSGSWGA 219
Qy 190 GGCTCAAACTTTTGGCGGTTGTTGGCAAGAGTAGTACCAACCGGGCAAGAATTGAC 249
Db 220 MSWKYMMRMRMRGNATGAGMKAMRASCMWRKRYAGSKTSYKSMWMCWTRSMKYCYTKA 279
Qy 250 CAGACAGAGATTATACCTTGATATATGATCAAGCGGCGAGTCCAGATGCCAGT 309
Db 280 RWTGYCYRKGWGRGRWYASKKYMMWRMWCWABMYRYSTGRASMMWRMYTMM 339
Qy 310 ATTGCAAGTGTCTATGTAATAGTGGATGATTATCCA--GAAAGTTCTGTAATA 367
Db 340 KWKYAMAPAAWMMWMMWMMRRACAAAATATATATTATGTAACAATCTGTACTT 399
Qy 368 AAGCAATATTAACAGATGTAAGTACTCAAAAAACGCAATTGTAGTGAGAGACAGT 424
Db 400 TAGCAATCTGGAGTTAGTTCAATAGTCAAAAGTCAGTTAATATTCTTAGAGGAAAGT 456

```

Search completed: March 16, 2004, 04:37:01
 Job time : 53.6647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-4
Perfect score: 456
Sequence: 1 atgaataaacatgttctatc.....ctgcgtgacacacgttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243257 seqs, 1840798884 residues
Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- Published Applications NA:*
- 1: /cgn2_6/ptodaca/2/pubpna/US07_PUBCOMB.seq:*
 - 2: /cgn2_6/ptodaca/2/pubpna/US06_NEW_PUB.seq:*
 - 3: /cgn2_6/ptodaca/2/pubpna/US06_NEW_PUB.seq:*
 - 4: /cgn2_6/ptodaca/2/pubpna/US06_PUBCOMB.seq:*
 - 5: /cgn2_6/ptodaca/2/pubpna/US07_NEW_PUB.seq:*
 - 6: /cgn2_6/ptodaca/2/pubpna/US07_PUBCOMB.seq:*
 - 7: /cgn2_6/ptodaca/2/pubpna/US08_NEW_PUB.seq:*
 - 8: /cgn2_6/ptodaca/2/pubpna/US08_PUBCOMB.seq:*
 - 9: /cgn2_6/ptodaca/2/pubpna/US09_PUBCOMB.seq:*
 - 10: /cgn2_6/ptodaca/2/pubpna/US09_PUBCOMB.seq:*
 - 11: /cgn2_6/ptodaca/2/pubpna/US09C_PUBCOMB.seq:*
 - 12: /cgn2_6/ptodaca/2/pubpna/US09C_PUBCOMB.seq:*
 - 13: /cgn2_6/ptodaca/2/pubpna/US10_PUBCOMB.seq:*
 - 14: /cgn2_6/ptodaca/2/pubpna/US10_PUBCOMB.seq:*
 - 15: /cgn2_6/ptodaca/2/pubpna/US10C_PUBCOMB.seq:*
 - 16: /cgn2_6/ptodaca/2/pubpna/US10_NEW_PUB.seq:*
 - 17: /cgn2_6/ptodaca/2/pubpna/US60_NEW_PUB.seq:*
 - 18: /cgn2_6/ptodaca/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	7.9	413	14 US-10-184-644-156 Sequence 156, App
2	36	7.9	413	14 US-10-184-644-156 Sequence 156, App
3	34.4	7.5	5184	14 US-10-247-671-55 Sequence 55, App1
4	33.8	7.4	1093	12 US-10-424-599-43422 Sequence 43422, A
5	33.8	7.4	1149	9 US-09-912-020-139 Sequence 139, App
6	33.8	7.4	1149	12 US-10-282-122A-6441 Sequence 6441, App
7	33.6	7.3	1299	12 US-10-282-122A-36109 Sequence 36109, A
8	33.2	7.3	1371	12 US-10-282-122A-17259 Sequence 17259, A
9	33	7.2	312	12 US-10-424-599-104721 Sequence 104721, A
10	33	7.2	1194	15 US-10-369-493-32803 Sequence 32803, A
11	33	7.2	1269	12 US-10-282-122A-35991 Sequence 35991, A
12	33	7.2	1837	12 US-10-282-122A-35601 Sequence 35601, A
13	33	7.2	319630	15 US-10-398-221-7 Sequence 7, App1
14	33	7.2	3011208	15 US-10-398-221-2058 Sequence 2058, App
15	32.8	7.2	746	15 US-10-027-632-14807 Sequence 14807, A

c	16	32.8	7.2	746	15	US-10-027-632-14808	Sequence 14808, A
c	17	32.8	7.2	2552	9	US-10-027-632-111939	Sequence 111939, A
c	18	32.6	7.1	351	9	US-09-983-965-4342	Sequence 4342, App
c	19	32.6	7.1	2000	15	US-10-260-238-1758	Sequence 1758, App
c	20	32.6	7.1	2250	15	US-10-369-493-46046	Sequence 46046, A
c	21	32.6	7.1	17419	14	US-10-239-676-100	Sequence 100, App
c	22	32.6	7.1	17419	14	US-10-311-455-1268	Sequence 1268, App
c	23	32.6	7.1	17419	14	US-10-240-453-112	Sequence 112, App
c	24	32.4	7.1	1320	12	US-10-282-122A-10004	Sequence 10004, A
c	25	32.4	7.1	2379	12	US-10-282-122A-9449	Sequence 9449, App
c	26	32.2	7.0	2634	12	US-10-282-122A-35828	Sequence 35828, A
c	27	32	7.0	2565	15	US-10-369-493-25168	Sequence 25168, A
c	28	31.8	7.0	409	9	US-09-864-761-1955	Sequence 2955, App
c	29	31.8	7.0	438	9	US-09-864-761-19736	Sequence 19736, A
c	30	31.8	7.0	708	14	US-10-184-644-584	Sequence 584, App
c	31	31.8	7.0	708	14	US-10-184-634-584	Sequence 584, App
c	32	31.8	7.0	1860	12	US-10-282-122A-27280	Sequence 27280, A
c	33	31.8	7.0	11138	10	US-09-960-870-5	Sequence 5, App1
c	34	31.8	7.0	11138	10	US-09-960-858-5	Sequence 5, App1
c	35	31.8	7.0	580073	14	US-10-205-220-1	Sequence 1, App1
c	36	31.8	7.0	640681	9	US-09-790-988-1	Sequence 1, App1
c	37	31.6	6.9	480	9	US-09-815-242-4053	Sequence 4053, App
c	38	31.6	6.9	480	12	US-10-282-122A-7419	Sequence 7419, App
c	39	31.6	6.9	575	15	US-10-027-632-222766	Sequence 222766, App1
c	40	31.6	6.9	2565	15	US-10-321-802-19	Sequence 19, App1
c	41	31.4	6.9	867	15	US-10-027-632-167010	Sequence 167010, App
c	42	31.4	6.9	867	15	US-10-027-632-167011	Sequence 167011, App
c	43	31.4	6.9	867	15	US-10-027-632-167012	Sequence 167012, App
c	44	31.4	6.9	1407	12	US-10-424-599-25067	Sequence 25067, A
c	45	31.4	6.9	2785	8	US-08-781-986A-133	Sequence 133, App

ALIGNMENTS

US-10-184-644-156	RESULT 1
Sequence 156, Application US/10184644	
Publication No. US20030044930A1	
GENERAL INFORMATION:	
APPLICANT: Baker, Kevin P.	
APPLICANT: Chen, Jian	
APPLICANT: Desnoyers, Luc	
APPLICANT: Goddard, Audrey	
APPLICANT: Godowski, Paul J.	
APPLICANT: Gurney, Austin L.	
APPLICANT: Pan, James	
APPLICANT: Smith, Victoria	
APPLICANT: Watanabe, Colin K.	
APPLICANT: Wood, William I.	
APPLICANT: Zhang, Zhenli	
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC	
FILE REFERENCE: P3430R1C227	
CURRENT APPLICATION NUMBER: US/10/184,644	
CURRENT FILING DATE: 2002-06-28	
Prior Application removed - See File Wrapper or Palm	
NUMBER OF SEQ ID NOS: 612	
SEQ ID NO 156	
LENGTH: 413	
TYPE: PRT	
ORGANISM: Homo Sapien	
US-10-184-644-156	
Query Match	7.9%; Score 36; DB 14; Length 413;
Best Local Similarity	11.6%; Pred. No. 2;
Matches	28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;
Oy	147 TGGTCAAGCTGGACATATATAGTGTCTACGCGAGGAGGCTCAAAATTTTGGC 206
Db	172 HDKDKVYKSGKRLRVGLTKRRNSGKRGSRGSRREASGQOREGTREHLEBRAGGR 231
Oy	207 GGTGTGGCCAGAGAGTAGTACCAACCGGCAAGATTGACCAAGAGATTATTA 266

```

Db      232  RRKSGRGRIAEGRSPFQWTRVKNTHIFKRWARGMDALDLDYLLLEKRAHKKKYM 291
Qy      267  CCTTGCATATATTGATGACGCGGAGTCCAGACGATATTTGGCAAGTGCTTA 326
Db      292  ELGISPTIKMFGGMHFGFNDADLVRFCSVDSBNLLYOYCDMESGSTSGYV 351
Qy      327  TGGTAATACGCGATGATTCACGAAAGTTCGTGTAATAAGCAATAATTACACAGTA 386
Db      352  LRKDPDKKMKRKIIAIVSGHWDVHGQKDVNVAVRITPLKVAQICMIHGNANCA 411
Qy      387  TG 388
Db      412  YG 413

```

RESULT 2

```

US-10-184-634-156
/ Sequence 156, Application US/10184634
/ Publication No. US20030068684A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Jian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Guirney, Austin L.
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P343081C217
/ CURRENT APPLICATION NUMBER: US/10/184,634
/ PRIOR FILING DATE: 2002-06-28
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO 156
/ LENGTH: 413
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-184-634-156

```

```

Query Match
Best Local Similarity 11.6%; Score 36; DB 14; Length 413;
Matches 28; Conservative 93; Mismatches 121; Indels 0; Gaps 0;

```

```

Qy      147  TGTCAAGTGGAGCTAATAATAGTCTAGTACGCGGAGGAGCTCAAACTTTGGC 206
Db      172  HDGKQYVSKKLRVGLMKMKNKSGKKRSGSKSRKREASGGDRGTREHIOERAKGR 231
Qy      207  GGTGTTCGCAAGAGTGTAGCAACCGGCAAGATTGACAGACGAGATTATTA 266
Db      232  RRKSGRGRIAEGRSPFQWTRVKNTHIFKRWARGMDALDLDYLLLEKRAHKKKYM 291
Qy      267  CCTTGCATATATTGATGACGCGGAGTCCAGACGATATTTGGCAAGTGCTTA 326
Db      292  ELGISPTIKMFGGMHFGFNDADLVRFCSVDSBNLLYOYCDMESGSTSGYV 351
Qy      327  TGGTAATACGCGATGATTCACGAAAGTTCGTGTAATAAGCAATAATTACACAGTA 386
Db      352  LRKDPDKKMKRKIIAIVSGHWDVHGQKDVNVAVRITPLKVAQICMIHGNANCA 411
Qy      387  TG 388
Db      412  YG 413

```

RESULT 3

```

US-10-247-671-55/c
/ Sequence 55, Application US/10247671

```

```

/ Publication No. US20030194721A1
/ GENERAL INFORMATION:
/ APPLICANT: Mikita, Thomas
/ APPLICANT: Shiffman, Dov
/ APPLICANT: Porter, Gordon, J.
/ APPLICANT: Kaser, Matthew R.
/ TITLE OF INVENTION: GENES EXPRESSED IN TREATED FOAM CELLS
/ FILE REFERENCE: PA-0050 US
/ CURRENT APPLICATION NUMBER: US/10/247,671
/ PRIOR FILING DATE: 2002-09-18
/ NUMBER OF SEQ ID NOS: 186
/ SOFTWARE: PERL Program
/ SEQ ID NO 55
/ LENGTH: 5184
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Incyte ID No. US20030194721A1 5185743CB1
US-10-247-671-55

```

```

Query Match
Best Local Similarity 7.5%; Score 34.4; DB 14; Length 5184;
Matches 74; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

```

Qy      51  GATTCACCCGACGAGGTATGATTGATTAATTCAGATATTAATTCGCGTAATGA 110
Db      2489  GCTTCACGCTCTGAATGATGATTTGTAAGCTGACCTACCTATTGGTGATGGAATA 2430
Qy      111  ATTGATGATGCTTATTAATGACGACGATTAATGCTGTAAGCTGGAAGCTAATATG 170
Db      2429  TTGATGATGCTGCTGCTGTAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2370
Qy      171  TGTCAAGTACGCGAGGAG 190
Db      2369  TGTGTGATGCTCTCACTGAG 2350

```

RESULT 4

```

US-10-424-599-43422/c
/ Sequence 43422, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J
/ APPLICANT: Kovalic, David K
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ PRIOR FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 43422
/ LENGTH: 1093
/ TYPE: DNA
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_139209C.1
US-10-424-599-43422

```

```

Query Match
Best Local Similarity 7.4%; Score 33.8; DB 12; Length 1093;
Matches 59; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```

```

Qy      294  TGCCACGATGCGGATATTGGCAAGTGTATGTAATGTCGATGATTAATCAAGA 353
Db      640  TGCAACAGTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581
Qy      354  AGTTCTGTAATAAACAATAATTACACAGTATGCTACT 394

```

Db 580 AGATCTGTGTAGTCTGACGATTACAGCTAATGTAATGTC 540

RESULT 5

US-09-912-020-139

Sequence 139, Application US/09912020

Patent No. US20020045592A1

GENERAL INFORMATION:

APPLICANT: Zykkind, Judith

APPLICANT: Ohlsen, Kari L.

APPLICANT: Trawick, John

APPLICANT: Forsyth, R. Allyn

APPLICANT: Froelich, Jamie M.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN

FILE REFERENCE: ELITRA.001DVI

CURRENT FILING DATE: 2001-07-23

PRIOR APPLICATION NUMBER: 09/492,709

PRIOR FILING DATE: 2000-01-27

PRIOR APPLICATION NUMBER: 60/117,405

PRIOR FILING DATE: 1999-01-27

NUMBER OF SEQ ID NOS: 485

SOFTWARE: PastSeq for Windows Version 3.0

SEQ ID NO 139

LENGTH: 1149

TYPE: DNA

ORGANISM: E. Coli

US-09-912-020-139

Query Match 7.4%; Score 33.8; DB 9; Length 1149;
Best Local Similarity 56.9%; Pred. No. 16;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 259 GATTATTAACCTTGATATGATGAGCGGCGGCAAGATGCGAATTTTCGAA 318

Db 442 GATAATTAACCTGGAATTAATGAGTTCAGGTGCGTAACACCCACGAGGTAATACATCGTCT 501

Qy 319 GGTGCTTATGCTTAATGATGCGATGATTATCCAGAAAGTCTGTGAATA 367

Db 502 GGCACACAGTGGTTACAGTCTCTTAATATGCTGAGAGCTTAATGTAATA 550

RESULT 6

US-10-282-122A-6441

Sequence 6441, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File wrapper or PALM.

NUMBER OF SEQ ID NOS: 78614

SOFTWARE: PatentIn version 3.1

SEQ ID NO 6441

LENGTH: 1149

TYPE: DNA

ORGANISM: Escherichia coli

US-10-282-122A-6441

Query Match 7.4%; Score 33.8; DB 12; Length 1149;
Best Local Similarity 56.9%; Pred. No. 16;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 259 GATTATTAACCTTGATATGATGAGCGGCGGCAAGATGCGAATTTTCGAA 318

Db 442 GATAATTAACCTGGAATTAATGAGTTCAGGTGCGTAACACCCACGAGGTAATACATCGTCT 501

Qy 319 GGTGCTTATGCTTAATGATGCGATGATTATCCAGAAAGTCTGTGAATA 367

Db 502 GGCACACAGTGGTTACAGTCTCTTAATATGCTGAGAGCTTAATGTAATA 550

RESULT 7

US-10-282-122A-36109

Sequence 36109, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zykkind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
FILE REFERENCE: 38-10(52052)B
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US/10/369,493
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 32803
LENGTH: 1194
TYPE: DNA
ORGANISM: Chloroflexus aurantiacus
US-10-369-493-32803

Query Match 7.2%; Score 33; DB 15; Length 1194;
Best Local Similarity 48.6%; Pred. No. 28;
Matches 90; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 167 ATAGTCTCAGTTACGGCAGGAGGCTCAAACTTTGGCGGTTGTGCGCAAGAAGTA 226
DB 737 ACACGTCTCGGTTGCACGAGGAAATCACCAATTCAGCCCTTGGCAAGCAAGAGGTG 678
QY 227 GTACCAACGGGCAAGATTGACGACAGAGATTATACCTTGCAATATATGATCAG 286
DB 677 TTGGCATAACCAACCAATGAGCATGATGACCTTACCAAGAGATTTTGCATCAG 618
QY 287 CGGCGACGCGCAAGATCCGATTTTGGCAAGGCTTATGATATGCGATGATTA 346
DB 617 CCGATCTTGTCAGTATTGTCGACCAACGATGAGGTTAGCAATGACGCTGCGGATTA 558
QY 347 TCCAG 351
DB 557 TTCAG 553

RESULT 11

US-10-282-122A-35991
Sequence 35991, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykend, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2001-02-16

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 35991
LENGTH: 1269
TYPE: DNA
ORGANISM: Streptococcus mutans
US-10-282-122A-35991

Query Match 7.2%; Score 33; DB 12; Length 1269;
Best Local Similarity 49.2%; Pred. No. 29;
Matches 87; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 236 GGGCAAGATTGACGACGAGAGATTATTAACCTTGCAATATATGATCAGCGGCACTG 295
DB 413 GAGCAAGATTCAACAGACGAGGTGTATATTAAGCCACTGCCATGCTTTAAAGGTG 472
QY 296 CCAAGATGCCGATTTTGGCAGGCTTATGATTAATGCGATGATTAATCCAGAAAG 355
DB 473 CTAACATTACATGATTTTCCAGTGTGGCGGACTCAGAACTTATGATGCTGCTA 532
QY 356 GTTCTGTAATAAGCAATATTACAGATGATGTAATCAAAAAAGGCAATTTAG 412
DB 533 CTTAGCTGATGAAACGATTAATGCAAAATGACGCTGAACAGAAATCGTTG 589

RESULT 12

US-10-282-122A-35601
Sequence 35601, Application US/10282122A
Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zykend, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614

Best Local Similarity 50.6%; Pred. No. 25;
Matches 79; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

Oy	239	CAAGATTGACCAACAGAGATTATTAACCTTGATATATTGATCAGCGGCAAGTCCA	298
Db	746	CAGAGACTTTCACCACTGTATATAAATGATTCATGATAGCTTGCGCAGCCCC	687
Oy	299	ACGATGCCAGTATTTGCAAGTGCTTATGTAATACTGCGATGATTTATCCAGAAAGTT	358
Db	686	ATGAGGCCGTGATGACTTAAGACGACTTGTAGAACTGCCATTTCAATGGCTAGAGTTT	627
Oy	359	CTGGTAATPAAGCAAAATATTACACAGTATGTAATC	394
Db	626	TAATTAATTTTGATCTTTGAAAAATAGTGAATC	591

Search completed: March 17, 2004, 08:15:26
Job time : 408.736 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 seconds

(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaacaacatcgtatc.....ttcgctgcacacagcttaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
EST:
1: em_estba:*
2: em_estbm:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hrc:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_hrc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gsa_hum:*
18: em_gsa_inv:*
19: em_gsa_pln:*
20: em_gsa_vrc:*
21: em_gsa_fun:*
22: em_gsa_mam:*
23: em_gsa_mus:*
24: em_gsa_pro:*
25: em_gsa_rtd:*
26: em_gsa_rhg:*
27: em_gsa_vr1:*
28: gb_gsa1:*
29: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	205.8	45.1	220	12	BI941542
2	56.8	12.5	207	9	AL368352
3	47	10.3	680	12	BU618688
4	40.2	8.8	956	29	CNS04QW8

5	37.8	8.3	533	10	BB703456
6	37	8.1	562	14	CP068694
7	37	8.1	671	12	BI311207
8	37	8.1	756	12	BM815787
9	36.6	8.0	957	13	BX453027
10	36.4	8.0	364	12	BM540455
11	36.4	8.0	597	28	AZ169114
12	36.4	8.0	657	10	BB449135
13	36.4	8.0	769	12	BG354883
14	36.2	7.9	491	28	AZ811111
15	36.2	7.9	532	28	AZ811109
16	36	7.9	678	12	BI452381
17	36	7.9	818	9	AL716706
18	35.6	7.8	615	28	AQ308392
19	35.6	7.8	624	28	AQ779365
20	35.4	7.8	705	12	BG125943
21	35.2	7.7	563	14	CA855727
22	35.2	7.7	744	12	BJ433259
23	35.2	7.7	1101	29	CNS0039C
24	35.2	7.7	1134	9	AL551649
25	35	7.7	491	28	BH068121
26	34.8	7.6	774	28	BZ527786
27	34.8	7.6	788	14	CA766697
28	34.8	7.6	823	28	BZ527794
29	34.8	7.6	992	13	BU180244
30	34.6	7.6	388	28	B86074
31	34.6	7.6	572	13	BX552747
32	34.6	7.6	1200	13	BX399167
33	34.4	7.5	412	12	BM167028
34	34.4	7.5	413	9	AL698409
35	34.4	7.5	595	12	BM852853
36	34.4	7.5	849	29	CG950661
37	34.4	7.5	862	29	CNS0295R
38	34.4	7.5	876	28	BZ794472
39	34.4	7.5	1050	13	BX349695
40	34.4	7.5	1201	13	BX324630
41	34.4	7.5	4397	29	AY413040
42	34.2	7.5	528	10	BE599976
43	34.2	7.5	528	28	AQ060604
44	34.2	7.5	531	12	BM323404
45	34.2	7.5	623	29	CC816176

ALIGNMENTS

RESULT 1
BI941542
LOCUS
DEFINITION
BI941542 220 bp mRNA linear EST 19-FEB-2003
dga26906.y1 Xenopus laevis gastrula non normalized Xenopus laevis
CDNA clone IMAGE:3749963 5' similar to SW:CSG_ECOLI P39828 MINOR
CURLIN SUBUNIT PRECURSOR, mRNA sequence.

ACCESSION
BI941542
VERSION
KEYWORDS
SOURCE
ORGANISM

Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
Xenopodinae; Xenopus.

REFERENCE
AUTHORS
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L.,
Pape,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y.,
Peterson,B., Gibbons,M., Harley,N., Rittler,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

TITLE
JOURNAL
COMMENT

Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
DNA Sequencing by: Washington Song
Source lab clone id: 319363
Trace considered overall poor quality This clone is available
royalty-free through LBNL; contact the IMAGE Consortium
(image@image.lbl.gov) for further information.
Seq primer: -40RP from gibco
High quality sequence stored

FEATURES

Smith College Department of Biological Sciences
Department of Biological Sciences
College, Northampton, MA, 01063, USA
Tel: 413/5853526
Fax: 413/5853786
Email: genome@smith.edu
Seq primer: Bluescript SK.
Location/Qualifiers
1..207

Source

1. 220
/organism="Xenopus laevis"
/mol_type="rRNA"
/db_xref="taxon:8355"
/clone="IMAGE:3749963"
/tissue="muscle"
/lab_host="gastula" (stages 10.5, 11.5 mixed)
/clone_lib="Xenopus laevis gastrula non normalized"
/note="vector: pBluescript SK-; Site 1: EcoRI; Site 2:
parts from stage 10.5 and stage 11.5 gastrulae."
EcoRI-XhoI cut cDNA was then ligated into Unilap-XR
(Stratagene) cut with EcoRI at the 5' end and XhoI at the 3'
Top10F. Clones were mass excised and used to infect
litter 15 g tryptone, 10g yeast extract, 5g NaCl, 36 mM
KH2PO4, 13.2 mM KH2PO4, 1.7 mM Na-citrate, 0.4 mM MgSO4
H2O, 6.8 mM (NH4)2SO4, 4 % w/v glycerol) and grown for 2
hours. Original library construction by Bruce Blumberg
(Cho et al 1991 Cell 67, 1111-1120)."

ORIGIN

and 1, filarial nematode parasite of humans, mRNA was prepared from approximately 200,000 microfilariae isolated from the skin of infected individuals from Kunba, Cameroon and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H free DNA pol. I. The library was 7.8 x 10⁶ independent recombinants and the average insert size is approximately 1kb. The library was constructed by Michelle Litvinchuk and Dr S.A. Williams. The library is available from email: genome@smith.edu."

ORIGIN

Query Match	45.14;	Score 205.8;	DB 12;	Length 220;
Best Local Similarity	96.84;	Pred. No. 4.7e-51;		
Matches 210;	Conservative			

Qy 80 CTAATTCAGATATATTACTTCGGCGGTAAATGAATTGAGTAAGTCTCTTATTAATCAGGAG 139
Db 3 CTAATTCATATATATATATCTTCGGCGGCACATGATGAGTAAGTCTCTTATTAATCAGGAG 62
Qy 140 CCATTAATTGTCAGAGCTGGGACTTAATTAATATAGTCTCACTTACGCAAGGAGGCTCAAAA 199
Db 63 CCATTAATTGTCAGAGCTGGGACTTAATTAATATAGTCTCACTTACGCAAGGAGGCTCAAAA 122
Qy 200 TTTTGGCGGTGTGTGGCCAGAGAGGTATGACCAACGGGCGAAAGATTGACCGACAGAG 259
Db 123 TTTTGGCGGTGTGTGGCCAGAGAGGTATGACCAACGGGCGAAAGATTGACCGACAGAG 182
Qy 260 ATTATTAACCTTCATATATATTATGATCAGGCGGCGAGTGC 296
Db 183 ATTATTAACCTTCATATATATTATGATCAGGCGGCGAGTGC 219

[illegible]

Query Match	12.5%	Score 56.8;	DB 9;	Length 207;
Best Local Similarity	95.1%	Pred. NO. 3.9e-06;		
Matches	58;	Conservative 0;	Mismatches 3;	Indels 0;
Gaps				
1	ATGAAACAAATTGTATTATGATGCTTAAACATATCTGGTGGCCCTGGAGATTGCAGCC	60		
Db	61	ATTAAACCAATTCATTCATTATGATGCTTAAACATACCTGGTGGCCCTGGAGATTGCAGCC	2	
Gy	61	G	61	
Db	1	G	1	

RESULT 3	B618688/c			
LOCUS	B618688	680 bp	mRNA	EST 01-OCT-2003
DEFINITION	B618688	Xenopus laevis Mochii normalized cDNA clone Xl186b22 5'		early gastrula library
ACCESSION	B618688			
VERSION	B618688.1	GI:37256713		
KEYWORDS	EST.			
SOURCE	Xenopus laevis (African clawed frog)			
ORGANISM	Xenopus laevis			
TITLE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
JOURNAL	Ampipibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;			
REFERENCE	Xenopodinae; Xenopus.			
AUTHORS	1 (Baes 1 to 680) Kikayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and Kohara,Y.			
COMMENT	Expressed genes in X. laevis embryo Unpublished (2001) Contact: Tadatsu Shin-i			

FEATURES

source	ncbi:csn:genes.nig.ac.jp
	The information of this clone is available through the following URL.
	http://xenopus.nibb.ac.jp .
	location/Qualifiers
	1..680

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186D2"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochi normalized Xenopus early gastrula library"

ORIGIN

Query Match 10.3%; Score 47; DB 12; Length 680;
Best Local Similarity 100.0%; Pred. No. 0.005;
Matches 47; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 410 TAGTGCAGACAGATCGCAATGCTATTCGGCTGACACACGTTAA 456
|||||
Db 680 TAGTGCAGACAGATCGCAATGCTATTCGGCTGACACACGTTAA 634

RESULT 4

LOCUS CNG04QM8 956 bp DNA linear GSS 01-SEP-2000
DEFINITION Tetradon nigroviridis genome survey sequence PUC-ORI end of clone 123D18 of library G from Tetradon nigroviridis, genomic survey sequence.
AL302777 GI:8183119

ACCESSION AL302777.1 GI:8183119
KEYWORDS GSS; genome survey sequence.
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE

1 Roest Croollius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C., Bernot, A., Filames, C., Winkler, P., Brotlier, P., Quetier, F., Saurin, M. and Weissenbach, J.
Estimate of human gene number provided by genome-wide analysis using Tetradon nigroviridis DNA sequence

TITLE

JOURNAL Nat. Genet. 25 (2), 235-238 (2000)

MEDLINE

PUBMED

20296633

REFERENCE

AUTHORS

2 Roest Croollius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C., Filames, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F., Saurin, M., Bernot, A. and Weissenbach, J.
Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigroviridis

TITLE

JOURNAL Genome Res. 10 (7), 939-949 (2000)

MEDLINE

PUBMED

20359837

REFERENCE

AUTHORS

3 (bases 1 to 956)
Genoscope.
Direct Submission
Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 Evry cedex - FRANCE (E-mail : seqef@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
This sequence is a single read and was generated as part of a large scale clone and sequencing project of the Tetradon nigroviridis genome. For more information, please take a look at <http://www.genoscope.cns.fr/Tetradon>.

FEATURES

SOURCE

1. 956
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="123D18"
/clone_lib="G"
/note="Genoscope sequence ID : C0BGI29D809SP1-end : PUC-ORI"

ORIGIN

Query Match 8.8%; Score 40.2; DB 29; Length 956;
Best Local Similarity 46.6%; Pred. No. 0.62;
Matches 129; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY 113 TGAGTAAGCTCTTCAATTTATCAGGACGACCAATATTTGTCAGCTGGACTATATATAGT 172
|||||
Db 430 TCAGTACGTGAGGACCACTATATCATCATGATGACATGACGAGGGCTACTATCAT 371

ORIGIN

TITLE

MEDLINE

PUBMED

20359837

REFERENCE

173 CTCAGTTACGGCAGGAGGCTCAAACTTTGGCGGTGTGGCCAGAGGTAGTACCA 232
|||||
Db 370 CACTGATCAGTCAGTCAGGAGCCACTATATCATCATGATGACATGACGAGGTACACTA 311

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

AUTHORS

233 ACCGGCAAGATTGACAGAGAGTTATTAACCTTGATATATATGATGAGGGGCA 292
|||||
QY 310 ATCAATCATCTATCATGATGATGAGAGTCACTATATCATCTGATCATGATGAGGG 251

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

293 GTGGCAACGATGCAAGTATTTCCGAAAGTGTATGATATATGATGATATATCAGA 352
|||||
Db 250 TCACATATCATCATCTATGATGATGATGATGATGATGATGATGATGATGATGATG 191

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

QY 353 AAGTTCTGTATTAAGCAATATTTACACATATG 389
|||||
Db 190 CAGGAGCCACTATATCATCTGATGATGATGATGATGATGATGATGATGATGATGATG 154

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

RESULT 5
LOCUS BB703456 533 bp mRNA linear EST 11-OCT-2001
DEFINITION BB703456 RIKEN full-length enriched, in vitro fertilized eggs Mus musculus cDNA clone 7420452F19 3', mRNA sequence.
BB703456
BB703456.1 GI:16052291
EST.

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 533)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)

ORIGIN

TITLE

MEDLINE

20359837

REFERENCE

wag1, K., Fujiwaka, S., Inoue, K., Togawa, Y., Iwawa, M., Ohara, E., Watanabe, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Kono, H., Fukunishi, Y., Shibata, K., Itoh, M., Carninci, P., Sugahara, Y. and Hayashizaki, Y.

1. .533
Location/Qualifiers

```

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="7420452F19"
/sex="female"
/risuse_type="in vitro fertilized eggs"
/dev_stage="egg"
/lab_host="DH10B"
/clone_lb="RIKEN full-length enriched, in vitro
fertilized eggs"

```

was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLX I. Cloning sites, 5' end: SalI, 3' end: BamHI¹¹

Query Match	8.3%;	Score 37.8;	DB 10;	Length 533;
Best Local Similarity	49.3%;	Pred. No. 2.8;		
Matches	99;	Conservative		

Qy	126	ATTTATATCAGCAGCCATAATTGTCACACTGGGACTATATATATGTCCTCAGTTACGGCA	185
Db	202	ATTTTTTATACCGGACATATGTTGCATGCTTTGAATCTTAACTTCAGAAAGCCAGAGGC	261
Qy	186	GGAGAGCTCAAACTTTTGGCGGTGTCGCCAAGAGTAGTAGCACCGGGCAAAAGAT	245
Db	262	AGGTGAGAGATCTATATATACGAGGAGTTCTCTCACTAGCAGGGCTTCACTGGGAAATATCT	321
Qy	246	TGACCAAGACAGAGATTATTAACCTTGACATATATGATGACAGCGGGCAGTGCCACGATGC	305
Db	322	GGCCCCCCCCAAAAAATTGTTGAGTTGATTTATTTGAGGGCTTATGTAGACCGAGGC	381
Qy	306	CAGTATTTGGCAAGGTGCTTA	326
Db	382	TATCTTGAACTGTTGTCTA	402

RESULT 6	
CF068694	
LOCUS	
DEFINITION	CF068694 562 bp mRNA
ACCESSION	E57659415 MTUS Medicago truncatula cDNA clone MTUS-11C11, mRNA sequence.
VERSION	CF068694
KEYWORDS	CF068694.1 GI:33105353
SOURCE	BST.
ORGANISM	Medicago truncatula (barrel medic)
	Medicago truncatula
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosoids I; Fabales; Fabaceae; Papilionoideae; Trifolieae; Medicago.
REFERENCE	1 (bases 1 to 562)

AUTHORS	TITLE	JOURNAL	COMMENT
VandenBosch, K., Endre, G., Silverstein, K., Town, C.D., Van Aken, S., ...	Utterback, T., Cheung, F., and Fraser, C.M. The Medicago truncatula 6K unigene set: cDNA clones selected and re-arrayed from various libraries Unpublished (2002)		Contact: VandenBosch K Department of Plant Biology University of Minnesota 220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA Tel: 612 624 2755 Fax: 612 625 1738 Email: kvandenbosch.umn.edu TIGR sequence name: MTUSK357K Alias Clone name: pGDS104c More information is available at: www.medicago.org Seq primer: Skm0d (CTA GAA CTA GTG GAT CC). Location/Qualifiers 1. 562

Location/Qualifiers
1. .562

XhoI. cDNA was prepared from poly(A)⁺ site 1. EcoRI; Site 2: was directionally ligated into the XbaI-XhoI vector from Stratagene and packaged using Gigapack III Gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-Zap phage using Ex-assist helper phage and propagated in XLOI cells.

Query Match	8.1%;	Score 37;	DB 14;	Length 562;
Best Local Similarity	53.9%;	Pred. No. 4.9;		
Matches	76;	Conservative	0.0;	Matched

		aa	index	U; Gaps	0;
OY	242	AGATTGACCAGACAGAGATTAATACTTCATATATTTGATCAGGGCGGCAGTGCACAG	301		
Dδ	59	AGATTGAACAGAGTGATGTATATGTGTGATTACAATAAGCTTTTATTTGTTCCA	118		
OY	302	ANGCCAGTATTTGCCAAGGGTCTTATGTATATCTGCATATATCCAGAAGGTTTG	361		
Dδ	119	AATATGACGACAGCAAAAGTGATGANTGTATATATGCGATTTCTATCTTCAAATCAAA	178		
OY	362	GTAATAAAGCAAAATATATACAC	382		
Dδ	179	ATAATTATGTATTAATTAATAC	199		

RESULT 7	
Bj311207	
LOCUS	Bj311207
DEFINITION	EST512957 GESD Medicago truncatula cDNA clone pGESD10A6 5' end.
ACCESSION	Bj311207
VERSION	Bj311207
KEYWORDS	EST.
SOURCE	Medicago truncatula (barrel medic)
ORGANISM	Medicago truncatula

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (pages 1 to 672)
Grusak, M.A., Samuel, D.A., Town, C.D., Van Aken, S., Uterback, T.,
Cho, J. and Fraser, C.M.
Ests from developing reproductive tissues of Medicago truncatula
Unpublished (2001)
Contact: Michael A. Grusak
USDA/ARS Children's Nutrition Research Center

REFERENCE Trichinelidae: Trichinella.
1 (bases 1 to 769)
AUTHORS Daub J., Connolly, B., Garate, T. and Blaxter, M.L.
TITLE A survey of genes expressed in the muscle stage larvae of the
JOURNAL parasitic nematode Trichinella spiralis
COMMENT Unpublished (2001)
Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
PCR Primers
FORWARD: T7SEQ (GTAATACGACTCACTATAGG)
BACKWARD: M13 Forward (CGCCAGGGTTTCCAGTCACGAC)
Seq primer: T7SEQ (GTAATACGACTCACTATAGG).

FEATURES
source
1. 769
/organism="Trichinella spiralis"
/mol_type="mRNA"
/db_xref="taxon:6334"
/clone="MBTSM1A133"
/sex="mixed"
/dev_stage="muscle stage larvae"
/note="Trichinella spiralis muscle stage larvae (BC)"
Site 2: Noci (3'end); The infective L1 larva of
Trichinella spiralis is a nematode parasite of mammalian
skeletal muscle. The library was constructed using muscle
larvae of the T. spiralis isolate 1553, and was provided
by Dr Bernadette Connolly, University of Aberdeen."

ORIGIN

Query Match
Best Local Similarity 8.0%; Score 36.4; DB 12; Length 769;
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 239 CAAGATTGACGACGAGATTATTAACCTTGCATATTTGATGAGCGGCGAGTCCCA 238
DB 78 CAGTTATGAATTCGATTAACGAAATATATAGCGCTTCAAAATTGGTCATCGACTTCAA 237
QY 299 ACGATGCCAGTATTTGCAAGGCTTATGATGATATGCGATGATTATCCAGAAAGTT 358
DB 138 TCATGCAAGCATTTGATGAAGATTTTACCTATGACTTATGATGATGAGCTTTGTT 197
QY 359 CTGCTATTAAGCAATATTACAGATGCTACTCAAAAACGCAATTGTAG 412
DB 198 CTGATGATTAAGCAATATTGTTATGAAGTGAAGAAATCAATGATTTGAG 251

RESULT 14
AZ811111 491 bp DNA linear GSS 20-FEB-2001
LOCUS 2M0077A11F Mouse 10kb plasmid UGCGM library Mus musculus genomic
DEFINITION clone UGCG2M0077A11 F, genomic survey sequence.
ACCESSION AZ811111
VERSION AZ811111.1 GI:12979042
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 491)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Irlam, H., Longacre, S., Mahmoud, M., Meenen, E., Petersen, T.,
Niederhuesern, A., Rose, R., Rose, R., Stokes, R., Tingey, A., von
Mouise whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL Unpublished (2000)
COMMENT Contact: Robert B. Weiss

University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddun@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0077 row: A column: 11
Seq primer: CGTTGTAAGACGACGCGCACT
Class: plasmid ends
High quality sequence stop: 491.
Location/Qualifiers
1. 491
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UGCG2M0077A11"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, Ti-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match
Best Local Similarity 7.9%; Score 36.2; DB 28; Length 491;
Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 270 TCATATATATGATGAGCGGCGAGTCCAGATGCTATTTGCAAGTGCTTATG 329
DB 110 TGAGATTAATATTTTCAGAGCTGTGTCAGCAAGGCAATATATGCTTACGTTAG 169
QY 330 TAATACGCCAGTATTTCCAGAAAGTCTGTAATAAGCAATATTACAGCTATG 389
DB 170 TACATGTCAGTATGTTCTGCAATATGATGTGTATACAGAGAAAGAAAAAGAA 229
QY 390 TACTCAAAAACGCAATTTGTGTGCAAGACA 422
DB 230 AAAAAAGAAAAAGAAAAAGAAATGAAAAAGAA 262

RESULT 15
AZ811109 532 bp DNA linear GSS 20-FEB-2001
LOCUS 2M0077A09F Mouse 10kb plasmid UGCGM library Mus musculus genomic
DEFINITION clone UGCG2M0077A09 F, genomic survey sequence.
ACCESSION AZ811109
VERSION AZ811109.1 GI:12979038
KEYWORDS GSS.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 532)

AUTHORS

Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
Rellly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
Niedermauern,A. and Wright,D.,Weiss,R.

TITLE

Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

JOURNAL

Unpublished (2000)

COMMENT

Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLc, UT
84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0077 row: A column: 09

Seq primer: CTTGTAAACGACGCGCCAGT

Class: plasmid ends

High quality sequence stop: 532.

FEATURES

Source

1..532

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="UUC2M0077A09"

/sex="Male"

/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"

/clone_id="Mouse 10kb plasmid UUC2M library"

/note="Vector: PMD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PMD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

ORIGIN

Query Match 7.9%; Score 36.2; DB 28; Length 532;
Best Local Similarity 52.3%; Pred. No. 8.4;

Matches 80; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

```

QY 270 TGCATATTATTCAGCGCGGCGAGCCAGATCCGATTTCCGCAAGTGCTTATGG 329
    |||||
DB 92 TGAGGATTAATATTTCAGAGCTGTGTCAAGCAAGCGATATCGTTTCATGTTAGAG 151
    |||||
QY 330 TAATACGCGATGATTCAGAAAGTTCTGTATTAAGCAATATTACACAGTATGG 389
    |||||
DB 152 TACATGTCGATAGTTCGAAATATGATTTGATTAACAAGAAAGAAAAAGAA 211
    |||||
QY 390 TACTCAAAAAACGCAATGTTAGTCAGAGACA 422
    |||||
DB 212 AAAAAAGAAAAAGAAAAAGATAGAAAAAGAAA 244
    |||||

```

Search completed: March 16, 2004, 04:28:52
Job time : 2237.91 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 206.439 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48

Sequence: 1 tatgatcagctggtaccg.....cccatgaatgagcatgca 48

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sgs:*
12: gb_sy:*
13: gb_un:*
14: gb_vt:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sgs:*
28: em_un:*
29: em_vt:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pla:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_hcg_hum:*
39: em_hcg_mus:*
40: em_hcg_oth:*
41: em_hcg_oth:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29.4	61.2	1797	3	LEIMSP52A
2	29.4	61.2	1860	3	LIGP63GEN
3	29.4	61.2	2104	3	LDO495002
4	29.4	61.2	2106	3	LDO495006
5	29.4	61.2	2106	3	LDO495007
6	29.4	61.2	2107	3	LDO495005
7	29.4	61.2	2109	3	LDO495003
8	29.4	61.2	2109	3	LDO495004
9	29.4	61.2	2892	3	LEIGP63B
10	29.4	61.2	2966	3	LEIGP63B
11	29.4	61.2	3047	3	LEIGP63A
12	29.4	61.2	3047	3	LHU48798
13	29.4	61.2	3105	3	LEIGP6A
14	28.4	59.2	2621	3	LEIGP63C
15	27.8	57.9	2050	3	LTU495010
16	27.8	57.9	2058	3	LAES32831
17	27.8	57.9	2077	3	LTR495009
18	27.8	57.9	2090	3	LARS52830
19	27.8	57.9	2125	3	LTR495008
20	27.8	57.9	2161	3	LMGP63
21	27.8	57.9	2187	3	LMGP63C1
22	27.8	57.9	3229	3	AF039721
23	27.8	57.9	11523	1	AE014950
24	26.8	55.8	57112	2	AC019976
25	26.8	55.8	88031	2	AC004378
26	26.8	55.8	124700	3	AC005558
27	26.8	55.8	126807	9	HS391022
28	26.8	55.8	171774	3	AC009909
29	26.8	55.8	295134	3	AE003582
30	26.4	55.0	1800	3	LEIMSP54A
31	26.2	54.6	2046	3	LEIGP63E
32	25.8	53.8	238365	2	AC107269
33	24.6	51.3	1964	3	LIGP63
34	24.6	51.3	2544	6	AX568314
35	24.6	51.3	3636	6	AR218846
36	24.6	51.3	3636	6	BD003758
37	24.6	51.3	11495	1	AB007386
38	24.6	51.3	12092	1	AB008447
39	24.6	51.3	147006	2	SPNEU1904
40	24.6	51.3	349980	6	AX571761
41	24.6	51.3	349980	6	AX571762
42	24.4	50.8	2240	8	AY118173
43	24.2	50.4	144239	10	AL627347
44	24.2	50.4	182340	2	AC099639
45	24.2	50.4	223974	2	AC106343

ALIGNMENTS

RESULT 1
LOCUS LEIMSP52A 1797 bp DNA linear INV 28-NOV-1994
DEFINITION Leishmania donovani glycoprotein 63 (mgs2) gene, complete cds.
VERSION L19563.1 GI:308887
ACCESSION L19563.1
KEYWORDS glycoprotein 63, surface protein, virulence factor, zinc protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE
1 (bases 1 to 1797)
Roberts,S.C., Swihart,K.G., Agey,M.W., Ramamorthy,R., Wilson,M.E.
and Donelson,J.B.
Sequence diversity and organization of the msp gene family encoding

Pred. No. is the number of results predicted by chance to have a

JOURNAL Mol. Biochem. Parasitol. 62 (2), 157-171 (1993)
MEDLINE 94187792
PUBMED 8139613

COMMENT

Original source text: Leishmania donovani (sub_species chagasi)

FEATURES

stationary stage promastigote DNA.

Location/Qualifiers

1..1797

/organism="Leishmania donovani"

/mol_type="genomic DNA"

/sub_species="chagasi"

/db_xref="taxon:5661"

/dev_stage="stationary stage promastigote"

1..1797

/gene="msp2"

1..1797

/gene="msp2"

/note="propeptide"

/codon_start=1

/product="glycoprotein 63"

/protein_id="AA53688.1"

/db_xref="GI:308888"

/translation="MSVSSSTHRRSVARLVRLAAGAIVAVGTAAAMAHAGV

WGLRIAVSTEDTDPAYHCAVGVGHIKRLGVDTCTAADIITDEKRIIVKHLIPO

ALQHTERLKVROVDKWKVTVGMGDVCSDFKVPPIHTDGLSTDFWVAVSVSE

DYLAAMATGVSDGHPAVGVINIPAINASRYDOLVTRVTHEMAHALGFSDITFD

KMLNVKTRKRNAPVINSSTAVAKARQVCDTLEIEMEDGCGAGSHIKM

RNAQDELMAPAAAGYSALTMAIFODLGYOAPFSAEVMGRRNAGCAFLSEKME

CGIVAFENGSCAQNASKVMAVOAFVPSDARCTIDGAFRRKRTETVNTSYAGIMY

VLCDTATRTYVOVGGSGYACTPGIIVELSTVSSAFEBEGYITCPPYEVCCQNV

AAKQGNAAARGRPRAAATLALVALLVAL"

1..39

/gene="msp2"

97..1794

/gene="msp2"

/product="glycoprotein 63"

Query Match

Best Local Similarity 61.2%; Score 29.4; DB 3; Length 1797;

Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGGTTACCGGCTGTTACCATGAAATGACATGC 47

751 TACGACAGCTGTGACGCTGTCTCAACGACGATGGCGACGC 797

RESULT 2

LOCUS LIGP63GEN 1860 bp DNA linear INV 16-DEC-1996

DEFINITION L. infantum GP63 gene.

VERSION 283677.1 GI:1743284

KEYWORDS GP63 gene; protease.

SOURCE Leishmania infantum

ORGANISM Leishmania infantum

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Leishmania.

1 (bases 1 to 1860)

Ortiz, G., Martinez, P. and Segovia, M.

Characterization of GP63 of Leishmania infantum

Unpublished

2 (bases 1 to 1860)

Ortiz, G.

Direct Submission

Submitted (12-DEC-1996) Ortiz, G., Facultad de Medicina, Universidad

de Murcia, Genetica y Microbiologia, Campus de Espinardo, Murcia,

Murcia, SPAIN, 30100

Location/Qualifiers

1..1860

/organism="Leishmania infantum"

FEATURES

Source

CDS

/mol_type="genomic DNA"

/strain="cutaneous"

/db_xref="taxon:5671"

/dev_stage="promastigote"

15..1815

/function="protease"

/codon_start=1

/product="GP63"

/protein_id="CAB06018.1"

/db_xref="GI:1743285"

/translation="MSVSSSTHRRSVARLVRLAAGAIVAVGTAAAMAHAGV

WGLRIAVSTEDTDPAYHCAVGVGHIKRLGVDTCTAADIITDEKRIIVKHLIPO

ALQHTERLKVROVDKWKVTVGMGDVCSDFKVPPIHTDGLSTDFWVAVSVSE

DYLAAMATGVSDGHPAVGVINIPAINASRYDOLVTRVTHEMAHALGFSDITFD

KMLNVKTRKRNAPVINSSTAVAKARQVCDTLEIEMEDGCGAGSHIKM

RNAQDELMAPAAAGYSALTMAIFODLGYOAPFSAEVMGRRNAGCAFLSEKME

CGIVAFENGSCAQNASKVMAVOAFVPSDARCTIDGAFRRKRTETVNTSYAGIMY

VLCDTATRTYVOVGGSGYACTPGIIVELSTVSSAFEBEGYITCPPYEVCCQNV

AAKQGNAAARGRPRAAATLALVALLVAL"

1..39

/gene="msp2"

97..1794

/gene="msp2"

/product="glycoprotein 63"

Query Match

Best Local Similarity 76.6%; Score 29.4; DB 3; Length 1860;

Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

1 TATGATCAGCTGGTTACCGGCTGTTACCATGAAATGACATGC 47

766 TACGACAGCTGTGACGCTGTCTCAACGACGATGGCGACGC 812

RESULT 3

LOCUS LIDO495002 2104 bp DNA linear INV 22-JUL-2002

DEFINITION Leishmania donovani mspc gene for GP63, strain

WHOW/KE/0000/Neal-R1.

VERSION AU495002.1 GI:21954455

KEYWORDS GP63; major surface protease; mspc gene.

SOURCE Leishmania donovani

Leishmania donovani

Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;

Leishmania.

1 Mauricio, I.L., Stothard, J.R. and Miles, M.A.

Genetic diversity in the Leishmania donovani complex

Unpublished

2 (bases 1 to 2104)

Mauricio, I.L.

Direct Submission

Submitted (11-JUL-2002) Mauricio, I.L., Infectious and Tropical

Diseases, London School of Hygiene and Trop. Med., Keppel Street,

WC1E 7HT, UNITED KINGDOM

Location/Qualifiers

1..2104

/organism="Leishmania donovani"

/mol_type="genomic DNA"

/strain="WHOW/KE/0000/Neal-R1"

/db_xref="taxon:5661"

/country="Kenya"

15..1940

/gene="mspC"

15..1940

/gene="mspC"

/note="constitutive major surface protease"

/codon_start=1

/allele="mspcdA1"

/product="GP63"

/protein_id="CAB06018.1"

/db_xref="GI:21954456"

/db_xref="GOA:O8MNZ5"

/db_xref="SPTRMBL:O8MNZ5"

/translation="MSVDSSTHRSVAARLVRLAAGAATAVTAAGTAAMAAGAV
OHRCTIDHMAQROSVACHHTAPGAVSAVGLPYVTITDTPAAADPRRPSAPVPAAN
WGAIRIANTSTEDLDTSAVHCARVQORVNNHAGATCTAEDITLDEKDIIVKLIIPQ
ALQHTERLKVROVDKNNVDMVDEICGDFVPAHITEGSNTDFPMVAAYSEK
GVLMAATTCOVPSDGHPRVGVINSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
ILVTTQMNIRKQDFNVSVNSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
RNADELMAAPAAAGYSALTMAIFODLGFYQADFSKAEPMWRNACATLSEKOME
ONITKMPAMFCNVSVNVNCTPSRLMIGTCGICGVPSPSPRYMOTFNISLGGYSPPL
DYCPFVIGYDGSQCNODASLATGFGAANNVSDARCTIDGAFRPNRTAAGYAGLC
ANVRCDIARTITSVQVGSMDYVNTPELRVELSTVSSAFEEGGITCPPYVEVCQAN
VKGADPAGDSDSSAGDADRAAMQRMNDRMGLATAAVLLGMVLSLMAALVVVWL
LLITCPMWCCKRFGGLPT"
15..131
/gene="mspC"
/protein_id="G63"
/product="G63"

sig_peptide
mat_peptide

Query Match 61.2%; Score 29.4; DB 3; Length 2104;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTGTTACCCGTTGTTACCCATGAATGCGACATGC 47
Db 765 TACGACGAGCTGTGACCGGTCTCTCACGACGAGATGCGCAGCG 811

RESULT 4
LDO495006 2106 bp DNA linear INV 22-JUL-2002
Leishmania donovani mspC gene for GP63, strain MCAN/IC/1977/Bummi3.
ACCESSION AJ495006
VERSION AJ495006.1 GI:21954463
KEYWORDS GP63; major surface protease; mspC gene.
SOURCE Leishmania donovani
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2106)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM

FEATURES
source
location/Qualifiers
1..2106
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MCAN/IC/1977/Bummi3"
/db_xref="taxon:5661"
/country="Iraq"
14..1939
/gene="mspC"
14..1939
/note="constitutive major surface protease"
/gene="mspC"
/codon_start=1
/allele="mspCldg2"
/product="GP63"
/protein_id="CAD42815.1"
/db_xref="GI:21954463"
/db_xref="GOA:Q8MM48"
/db_xref="SPTREMBL:Q8MM48"
/translation="MSVDSSTHRSVAARLVRLAAGAATAVTAAGTAAMAAGAV
OHRCTIDHMAQROSVACHHTAPGAVSAVGLPYVTITDTPAAADPRRPSAPVPAAN
WGAIRIANTSTEDLDTSAVHCARVQORVNNHAGATCTAEDITLDEKDIIVKLIIPQ
ALQHTERLKVROVDKNNVDMVDEICGDFVPAHITEGSNTDFPMVAAYSEK
GVLMAATTCOVPSDGHPRVGVINSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
ILVTTQMNIRKQDFNVSVNSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
RNADELMAAPAAAGYSALTMAIFODLGFYQADFSKAEPMWRNACATLSEKOME
ONITKMPAMFCNVSVNVNCTPSRLMIGTCGICGVPSPSPRYMOTFNISLGGYSPPL
DYCPFVIGYDGSQCNODASLATGFGAANNVSDARCTIDGAFRPNRTAAGYAGLC
ANVRCDIARTITSVQVGSMDYVNTPELRVELSTVSSAFEEGGITCPPYVEVCQAN
VKGADPAGDSDSSAGDADRAAMQRMNDRMGLATAAVLLGMVLSLMAALVVVWL
LLITCPMWCCKRFGGLPT"

ILVTTQMNIRKQDFNVSVNSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
RNADELMAAPAAAGYSALTMAIFODLGFYQADFSKAEPMWRNACATLSEKOME
ONITKMPAMFCNVSVNVNCTPSRLMIGTCGICGVPSPSPRYMOTFNISLGGYSPPL
DYCPFVIGYDGSQCNODASLATGFGAANNVSDARCTIDGAFRPNRTAAGYAGLC
ANVRCDIARTITSVQVGSMDYVNTPELRVELSTVSSAFEEGGITCPPYVEVCQAN
VKGADPAGDSDSSAGDADRAAMQRMNDRMGLATAAVLLGMVLSLMAALVVVWL
LLITCPMWCCKRFGGLPT"
14..130
/gene="mspC"
131..1822
/gene="mspC"
/product="GP63"

sig_peptide
mat_peptide

Query Match 61.2%; Score 29.4; DB 3; Length 2106;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTGTTACCCGTTGTTACCCATGAATGCGACATGC 47
Db 764 TACGACGAGCTGTGACCGGTCTCTCACGACGAGATGCGCAGCG 810

RESULT 5
LDO495007 2106 bp DNA linear INV 22-JUL-2002
Leishmania donovani mspC gene for GP63, strain
MHOM/SU/1984/Marz-Krim.
ACCESSION AJ495007
VERSION AJ495007.1 GI:21954465
KEYWORDS GP63; major surface protease; mspC gene.
SOURCE Leishmania donovani
ORGANISM Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2106)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM

FEATURES
source
location/Qualifiers
1..2106
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/SU/1984/Marz-Krim"
/db_xref="taxon:5661"
14..1939
/gene="mspC"
14..1939
/note="constitutive major surface protease"
/codon_start=1
/allele="mspCldh"
/product="GP63"
/protein_id="CAD42816.1"
/db_xref="GI:21954466"
/db_xref="GOA:Q8MM48"
/db_xref="SPTREMBL:Q8MM48"
/translation="MSVDSSTHRSVAARLVRLAAGAATAVTAAGTAAMAAGAV
OHRCTIDHMAQROSVACHHTAPGAVSAVGLPYVTITDTPAAADPRRPSAPVPAAN
WGAIRIANTSTEDLDTSAVHCARVQORVNNHAGATCTAEDITLDEKDIIVKLIIPQ
ALQHTERLKVROVDKNNVDMVDEICGDFVPAHITEGSNTDFPMVAAYSEK
GVLMAATTCOVPSDGHPRVGVINSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
ILVTTQMNIRKQDFNVSVNSSTAVAKARQYCGDTLEVEIEDOGAGASGSHIKM
RNADELMAAPAAAGYSALTMAIFODLGFYQADFSKAEPMWRNACATLSEKOME
ONITKMPAMFCNVSVNVNCTPSRLMIGTCGICGVPSPSPRYMOTFNISLGGYSPPL
DYCPFVIGYDGSQCNODASLATGFGAANNVSDARCTIDGAFRPNRTAAGYAGLC
ANVRCDIARTITSVQVGSMDYVNTPELRVELSTVSSAFEEGGITCPPYVEVCQAN
VKGADPAGDSDSSAGDADRAAMQRMNDRMGLATAAVLLGMVLSLMAALVVVWL
LLITCPMWCCKRFGGLPT"

Query Match	61.2%	Score 29.4	DB 3	Length 2106
Best Local Similarity	76.6%	Fred. No. 2.8		
Matches	36	Mismatches	11	Indels 0; Gaps 0
1	TATGATCAGCTGTTACCGCGTGTGTTA		TACCATTAATGACACATGC	47
764	TACGACGAGCTGATGACACAGTGTCTACG		CCACGACAGATGGCGACGC	810

RESULT 6	LDO495005	2107 bp	DNA	linear	INV 22-JUL-2002
LOCUS	Leishmania donovani mspc gene for GP63, strain				
DEFINITION	MCAN/1Q/1981/Sukkar2.				
ACCESSION	AJ495005				
VERSION	AJ495005.1	GI:21954461			
KEYWORDS	GP63; major surface protease; mspc gene.				
SOURCE	Leishmania donovani				
ORGANISM	Leishmania donovani				
	Bukariyota; Euglenozoa; Kinetoplastida; Trypanosomatidae;				
	Leishmania.				
REFERENCE	1				
AUTHORS	Mauricio, I.L., Stothard, J.R. and Miles, M.A.				
TITLE	Genetic diversity in the Leishmania donovani complex				
JOURNAL	unpublished				
REFERENCE	2 (bases 1 to 2107)				
AUTHORS	Mauricio, I.L.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-JUL-2002)				
	Mauricio I.L., Infectious and Tropical				
	Diseases, London School of Hygiene and Trop. Med., Keppel Street,				
	WC1E 7HT, UNITED KINGDOM				
FEATURES	Location/Qualifiers				
source	1..2107				

sig_peptide

Query March	61.2%	Score 29.4	DB 3	Length 2107
Best Local Similarity	76.6%	Prod No. 2.8		
Matches	36	Conservative	0	Mismatches 11
				Indels 0
				Gaps 0
QY	1	TATGATCAGCTGGTTACCCGGTGTGTATCCATGAATGGCAATGCAATGTC	47	
Db	764	TACGACAGCTGTGACACCGTGTGTATCCAGCAAGATGGCCGACGC	810	

RESULT	7
LDO495003	
LOCUS	
DEFINITION	
ACCESSION	LDO495003
VERSION	2109 bp
KEYWORDS	DNA
SOURCE	linear
ORGANISM	INV 22-JUL-2002
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	
TITLE	
JOURNAL	
source	

LDO495003
Leishmania donovani mspc gene for GP63, strain
 MHOM/KE/1954/LNC-153.
 AU495003
 AU495003.1 GI:21954457
 GP63; major surface protease; mspc gene.
Leishmania donovani
Leishmania donovani
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Leishmania.
 Mauricio I.L., Stochard, V.R. and Miles, M.A.
 Genetic diversity in the *Leishmania donovani* complex
 2 (bases 1 to 2109)
 Mauricio, I.L.
 Direct Submission
 Submitted (11-JUN-2002) Maurício I.L., Infectious and Tropical
 Diseases, London School of Hygiene and Trop. Med., Keppel Street,
 WC1E 7HT, UNITED KINGDOM
 1..2109
 location/Qualifiers

```
sig_peptide
mat_peptide
ORIGIN
```

[illegible]

Query Match 61.2%; Score 29.4; DB 3; Length 2109;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGACATGC 47
765 TACGACACGCTGTGACCGCTGTGTCACGACGAGATGGCGACGC 811

RESULT 8
LDO495004 2109 bp DNA linear INV 22-JUL-2002
LOCUS Leishmania donovani mspc gene for GP63, strain
DEFINITION MHOM/CN/0000/Mang1el1.
ACCESSION AJ295004
VERSION AJ295004.1 GI:21954459
KEYWORDS GP63; major surface protease; mspc gene.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2109)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
SUBMITTED (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM
LOCATION/Qualifiers
1. 2109
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/strain="MHOM/CN/0000/Mang1el1"
/db_xref="taxon:5661"
/country="China"
15. 1940
/gene="mspC"
15. 1940
/note="constitutive major surface protease"
/codon_start=1
/allele="mspCldf"
/product="mspC"
/protein_id="CAD2813.1"
/db_xref="GI:21954460"
/db_xref="GOA:O8MN23"
/db_xref="SPRMBL:O8MN23"
/translation="MSYDSSSTRHRSVARRLYRLAAGAIVTAAGTAAMAAGAV
QHRCTIDAMQARVQSVARRHTAPGAVSAGLPVTLDTAAADRSGAPVVRAN
WGLRIRAVSTEDLTDPAYHCAVGOHVNNAAGATCTADILTDKRDILVXLIPO
ALQHTERLKVQVQDKMVTDMWDEICDPKVPQAHITGFSNTDFVMYVASPSE
GVLAAMTTCCVPSDGHPAVGVINIPAAINASRYDLPVTRVTHEMAHALGFSGFPE
ILVVTOMNIRGDFDPVTVNSGRAVAKARBOVCPTLLETIEDOGASASGHITK
RNAQDELMAAPAAAGTYSALTMAIFDULGTYQADFSAEVMPKRNAGCAFLEKCKE
RNTTKPAMFCNENEVMTWCPTSRILSGKGVGRHDPDPYQYFTDPSLAGISAFMD
DYCEPVIYGDGSCNODASLAAGFFSAFNVFSDARCIDGAFRKNKTANGYAGLC
ANVCDAITRTYSVQVHGSMDYVNTCTGLRVELSTVSFAFEGGYITCPYVEVCOGN
VOAKDPAAGDSSSSAGDAADRAAQRMDRMAAGLTAAMVLLGWLISLALVYVWL
LILTCPPWCKCFGLPFR"

15. 131
/gene="mspC"
132. 1823
/gene="mspC"
/product="GP63"

ORIGIN
Query Match 61.2%; Score 29.4; DB 3; Length 2109;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Query Match 61.2%; Score 29.4; DB 3; Length 2892;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGACATGC 47
765 TACGACACGCTGTGACCGCTGTGTCACGACGAGATGGCGACGC 811

RESULT 9
LEIGP63D 2892 bp mRNA linear INV 26-APR-1993
LOCUS Leishmania donovani chagasi surface protease (log glycoprotein 63,
DEFINITION GP63) mRNA, complete cds.
ACCESSION M80672
VERSION M80672.1 GI:159328
KEYWORDS log glycoprotein 63; surface protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.

REFERENCE 1 (bases 1 to 2892)
AUTHORS Ramamoorthy, R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C.
and Wilson, M.E.
TITLE Three distinct RNAs for the surface protease gp63 are
JOURNAL differentially expressed during development of Leishmania donovani
MEDLINE J. Biol. Chem. 267 (3), 1888-1895 (1992)
PUBMED 92112918
1370484
COMMENT Original source text: Leishmania donovani (sub-species chagasi)
CDNA to mRNA.

FEATURES
source Location/Qualifiers
1. 2892
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
1. 2892
/gene="GP63"
114. 1913
/gene="GP63"
/note="log glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AAA29238.1"
/db_xref="GI:159328"

1. 230
/translation="MSYDSSSTRHRSVARRLYRLAAGAIVTAAGTAAMAAGAV
QHRCTIDAMQARVQSVARRHTAPGAVSAGLPVTLDTAAADRSGAPVVRAN
WGLRIRAVSTEDLTDPAYHCAVGOHVNNAAGATCTADILTDKRDILVXLIPO
ALQHTERLKVQVQDKMVTDMWDEICDPKVPQAHITGFSNTDFVMYVASPSE
GVLAAMTTCCVPSDGHPAVGVINIPAAINASRYDLPVTRVTHEMAHALGFSGFPE
ARLIESISNVRHSDPVPVTVNSGRAVAKARBOVCPTLLETIEDOGASASGHITK
RNAQDELMAAPAAAGTYSALTMAIFDULGTYQADFSAEVMPKRNAGCAFLEKCKE
RNTTKPAMFCNENEVMTWCPTSRILSGKGVGRHDPDPYQYFTDPSLAGISAFMD
CCVVPYDGGSCAQRASEAGAPKGFNFSDARCIDGAFRKNKTANGYAGLC
ANVCDAITRTYSVQVHGSMDYVNTCTGLRVELSTVSFAFEGGYITCPYVEVCOGN
VOAKDGNAAAGRRGPRAAATALLVALAVAL"

114. 230
/gene="GP63"
405. 1910
/gene="GP63"
/product="surface protease"
/note="log glycoprotein 63"
2892
/gene="GP63"

ORIGIN
Query Match 61.2%; Score 29.4; DB 3; Length 2892;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGACATGC 47
864 TACGACACGCTGTGACCGCTGTGTCACGACGAGATGGCGACGC 910

RESULT 10
LOCUS LEIGP63B 2966 bp mRNA linear INV 26-APR-1993
DEFINITION Leishmania donovani chagasi surface protease (stationary
glycoprotein 63, gp63) mRNA, complete cds.
ACCESSION M80669.1 GI:159324
VERSION M80669.1
KEYWORDS stationary glycoprotein 63; surface protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 2966)
AUTHORS Ramamoorthy R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C.
TITLE Three distinct RNAs for the surface protease gp63 are
differentially expressed during development of Leishmania donovani
chagasi promastigotes to an infectious form
JOURNAL J. Biol. Chem. 267 (3), 1888-1895 (1992)
MEDLINE 92112918
PUBMED 1370484
COMMENT Original source text: Leishmania donovani (sub_species chagasi)
CDS to mRNA.
FEATURES
source
Location/Qualifiers
1..2966
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
1..2966
/gene="GP63"
129..1928
/gene="GP63"
/note="stationary glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AA029236.1"
/db_xref="GI:159325"
/translation="MSVDSSTHRRSVARLVRLAAGAATVAVGTAAAHAGAV
QHRCHDMMQARVRSVARHHTAPGASAVGLPYVTLDTAAADRPGSPATVRRAN
WGALRIAVSTEDLTPAHYHCAVGOHKKRLGGVDICTADILTDKRDILVHLLPQ
ALQHLTERLKYROYDDKWKVGMDDKSPKPPATITDGLSTNTPFVAVSPSEB
GVLAWAATCOVPSGHPAVGVINIPANINISRDQVTRVTHEMAHALGFSVFPFEG
ARIESISNVRRKDDPVDVINSSTVAKAABQYGCDDLEVLIEDGAGASGSHIKM
RNADDELMAAPAAAGYSALTMAIFQDLGYQADFSAEVMFGRAAGCAPISSEKME
RNITKWPAMFCNENEVTRCPTSRSLISGCGVTRHDPDPYMOYFTDPSLAGISAFMD
CCPVEEPYDGSCAQRASEAGAPFKGNVPSDARCTIDGAPFRKTSHGQIKSYAGICA
NVRCDTARTYTSVQVHGSGYANCTPGLRVELSTVSAPFEGYITCPPEVVCQGNV
QAADGMAAAGRGRPRAAATALLVALAVAL"
129..245
/gene="GP63"
420..1925
/gene="GP63"
/product="surface protease"
/note="stationary glycoprotein 63"
2966
/gene="GP63"
ORIGIN
Query Match 61.2%; Score 29.4; DB 3; Length 2966;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Db 1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGCAGTC 47
879 TAGACACAGCTGTGACGCTGTCGTCACGACGAGATGGCGCAGC 925
1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGCAGTC 47
879 TAGACACAGCTGTGACGCTGTCGTCACGACGAGATGGCGCAGC 925
LOCUS LEIGP63A 3047 bp DNA linear INV 26-APR-1993
DEFINITION L.chagasi major surface glycoprotein (gp63) gene, complete cds.

ACCESSION M28527
VERSION M28527.1 GI:159322
KEYWORDS glycoprotein; protease; surface antigen.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 3047)
AUTHORS Miller, R.A., Reed, S.G., and Parsons, M.
TITLE Leishmania gp63 molecule implicated in cellular adhesion lacks an
Arg-Gly-Asp sequence
JOURNAL Mol. Biochem. Parasitol. 39 (2), 267-274 (1990)
MEDLINE 9205976
PUBMED 2320059
COMMENT Original source text: L. chagasi (isolate MHOM/BR/82/BA-2C1a) DNA,
clones p1c63-[1 and 2].
Draft entry and computer-readable sequence for [1] kindly provided
by M. Parsons, 03-OCT-1989.
FEATURES
source
Location/Qualifiers
1..3047
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
496..2295
/note="gp63 protein"
/codon_start=1
/protein_id="AA029235.1"
/db_xref="GI:159323"
/translation="MSVDSSTHRRSVARLVRLAAGAATVAVGTAAAHAGAV
QHRCHDMMQARVRSVARHHTAPGASAVGLPYVTLDTAAADRPGSPATVRRAN
WGALRIAVSTEDLTPAHYHCAVGOHKKRLGGVDICTADILTDKRDILVHLLPQ
ALQHLTERLKYROYDDKWKVGMDDKSPKPPATITDGLSTNTPFVAVSPSEB
GVLAWAATCOVPSGHPAVGVINIPANINISRDQVTRVTHEMAHALGFSVFPFEG
ARIESISNVRRKDDPVDVINSSTVAKAABQYGCDDLEVLIEDGAGASGSHIKM
RNADDELMAAPAAAGYSALTMAIFQDLGYQADFSAEVMFGRAAGCAPISSEKME
RNITKWPAMFCNENEVTRCPTSRSLISGCGVTRHDPDPYMOYFTDPSLAGISAFMD
CCPVEEPYDGSCAQRASEAGAPFKGNVPSDARCTIDGAPFRKTSHGQIKSYAGICA
NVRCDTARTYTSVQVHGSGYANCTPGLRVELSTVSAPFEGYITCPPEVVCQGNV
QAADGMAAAGRGRPRAAATALLVALAVAL"
ORIGIN
Query Match 61.2%; Score 29.4; DB 3; Length 3047;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Db 1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGCAGTC 47
1246 TAGACACAGCTGTGACGCTGTCGTCACGACGAGATGGCGCAGC 1292
1 TATGATCAGCTGGTTACCCGTTGTTTACCCATGAATGGCAGTC 47
1246 TAGACACAGCTGTGACGCTGTCGTCACGACGAGATGGCGCAGC 1292
LOCUS LIU48798 3047 bp DNA linear INV 15-JAN-1998
DEFINITION Leishmania infantum surface glycoprotein gp63 (gp63) gene, complete
cds.
ACCESSION U48798
VERSION U48798.1 GI:1213329
KEYWORDS Leishmania infantum
SOURCE Leishmania infantum
ORGANISM Leishmania infantum
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 3047)
AUTHORS Gonzalez-Aseguinolaza, G., Almazan, F., Rodriguez, J.F., Marguet, A.
TITLE Cloning of the gp63 surface protease of Leishmania infantum.
Differential post-translational modifications correlated with
different infective forms
JOURNAL Biochim. Biophys. Acta 1361 (1), 92-102 (1997)
MEDLINE 97388259
PUBMED 9247093
REFERENCE 2 (bases 1 to 3047)

AUTHORS Gonzalez-Abeguinolaza, G., Almazan, F., Rodriguez-Aguirre, J.F.,
Marquet, A. and Larraga, V.
TITLE Direct Submission
JOURNML Submitted (08-FEB-1996) Vicente Larraga, Proteins Structure, Center
of Biological Research, Spanish Research Council, Velazquez 144,
Madrid 28006, Spain

FEATURES

SOURCE location/Qualifiers
1..3047
/organism="Leishmania infantum"
/mol_type="genomic DNA"
/db_xref="taxon:5671"
gene 496..2295
/gene="gp63"
496..2295
/gene="gp63"
/note="protease"
/codon_start=1
/product="surface glycoprotein gp63"
/protein_id="AAB96339.1"
/db_xref="GI:1213330"
/translation="MSVDSSTHRSVAAVLVRLAAGAATAVTAAGTAAMAAGAV
OHRCHIDMQRARVRSVARRHTAPGASAVGLPYVTLDTAAARRPESAPTVRAAN
WGALRIAVSTEDLTPAHCARVQDHIKRLGVDICTAEDILTDKRDILVKHLIQ
ALQHLTERLKVROVDKRWYMGDVCSDPKVPPIITDGLSNTDFVMVAVSPSE
GVLAATTCQVPSDGHPAVGVINIPANIASRYDQLVTRVTHMAALGFSVGFEG
ARLESISNVKHDPEVINSSTAFAKROGCDTLEYEIDOGAGSGSHIKM
RNADDELAPAAAGVYSALTMATFODLGFYQADPSKAEVMPKRNACALSKCHE
NRTKMPAMFCNENBYTRCTSLSLCKCGVTHPDLPPYWOYFTDLSLGFAMD
CSPVEVIGSDSCAQRASLGAEPKGFNFSDARCIDAGAPRPTSHQIIKSYAGLCV
NVRCDTARTYSVQVHGSGYANCTPGLRVELSTVSSAFEBGGYITCPYVEVCOGAV
QAAKDGNAAGRGRPAAATALLVALAVAL"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 3047;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Oy 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGC 47
Db 1246 TACACACGCTGTGACCGCTGTCTCTACGACGAGATGGCGACGC 1292

RESULT 13
LEIGPAA 3105 bp DNA linear INV 26-APR-1993
LOCUS L.donovani.
ACCESSION M60048
VERSION M60048.1 GI:159334
KEYWORDS glycoprotein 63.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 3105)
AUTHORS Webb, J.R., Button, L.L. and McMaster, W.R.
TITLE Heterogeneity of the genes encoding the major surface glycoprotein
of Leishmania donovani
JOURNAL Mol. Biochem. Parasitol. 48 (2), 173-184 (1991)
MEDLINE 92107220
PUBMED 1762629

COMMENT Original source text: L.donovani DNA.
FEATURES Location/Qualifiers
1..3105
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/db_xref="taxon:5661"
gene 101..1873
/gene="gp63"
101..1873
/gene="gp63"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AAA29244.1"

/db_xref="GI:159335"
/translation="MSVDSSTHRSVAAVLVRLAAGAATAVTAAGTAAMAAGAV
OHRCHIDMQRARVRSVARRHTAPGASAVGLPYVTLDTAAARRPESAPTVRAAN
WGALRIAVSTEDLTPAHCARVQDHIKRLGVDICTAEDILTDKRDILVKHLIQ
ALQHLTERLKVROVDKRWYMGDVCSDPKVPPIITDGLSNTDFVMVAVSPSE
GVLAATTCQVPSDGHPAVGVINIPANIASRYDQLVTRVTHMAALGFSVGFEG
ARLESISNVKHDPEVINSSTAFAKROGCDTLEYEIDOGAGSGSHIKM
RNADDELAPAAAGVYSALTMATFODLGFYQADPSKAEVMPKRNACALSKCHE
NRTKMPAMFCNENBYTRCTSLSLCKCGVTHPDLPPYWOYFTDLSLGFAMD
CSPVEVIGSDSCAQRASLGAEPKGFNFSDARCIDAGAPRPTSHQIIKSYAGLCV
NVRCDTARTYSVQVHGSGYANCTPGLRVELSTVSSAFEBGGYITCPYVEVCOGAV
QAAKDGNAAGRGRPAAATALLVALAVAL"

ORIGIN

Query Match 61.2%; Score 29.4; DB 3; Length 3105;
Best Local Similarity 76.6%; Pred. No. 2.8;
Matches 36; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Oy 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGC 47
Db 821 TACACACGCTGTGACCGCTGTCTCTACGACGAGATGGCGACGC 867

RESULT 14
LEIGP63C 2621 bp mRNA linear INV 26-APR-1993
LOCUS Leishmania donovani chagasi surface protease (constitutive
glycoprotein 63, gp63) mRNA, complete cds.
ACCESSION M80671
VERSION M80671.1 GI:159326
KEYWORDS constitutive glycoprotein 63; surface protease.
SOURCE Leishmania donovani
ORGANISM Leishmania donovani
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1 (bases 1 to 2621)
AUTHORS Ramamoorthy, R., Donelson, J.E., Paetz, K.E., Maybodi, M., Roberts, S.C.
and Wilson, M.E.
TITLE Three distinct RNAs for the surface protease gp63 are
differentially expressed during development of Leishmania donovani
JOURNAL J. Biol. Chem. 267 (3), 1888-1895 (1992)
MEDLINE 92112918
PUBMED 1370484

COMMENT Original source text: Leishmania donovani (sub-species chagasi)
CDNA to mRNA.
FEATURES Location/Qualifiers
1..2621
/organism="Leishmania donovani"
/mol_type="mRNA"
/sub_species="chagasi"
/db_xref="taxon:5661"
gene 1..2621
/gene="gp63"
132..2051
/gene="gp63"
/note="constitutive glycoprotein 63"
/codon_start=1
/product="surface protease"
/protein_id="AAA29237.1"
/db_xref="GI:159327"
/translation="MSVDSSTHRSVAAVLVRLAAGAATAVTAAGTAAMAAGAV
OHRCHIDMQRARVRSVARRHTAPGASAVGLPYVTLDTAAARRPESAPTVRAAN
WGALRIAVSTEDLTPAHCARVQDHIKRLGVDICTAEDILTDKRDILVKHLIQ
ALQHLTERLKVROVDKRWYMGDVCSDPKVPPIITDGLSNTDFVMVAVSPSE
GVLAATTCQVPSDGHPAVGVINIPANIASRYDQLVTRVTHMAALGFSVGFEG
ARLESISNVKHDPEVINSSTAFAKROGCDTLEYEIDOGAGSGSHIKM
RNADDELAPAAAGVYSALTMATFODLGFYQADPSKAEVMPKRNACALSKCHE
NRTKMPAMFCNENBYTRCTSLSLCKCGVTHPDLPPYWOYFTDLSLGFAMD
CSPVEVIGSDSCAQRASLGAEPKGFNFSDARCIDAGAPRPTSHQIIKSYAGLCV
NVRCDTARTYSVQVHGSGYANCTPGLRVELSTVSSAFEBGGYITCPYVEVCOGAV
QAAKDGNAAGRGRPAAATALLVALAVAL"

COMMENT Original source text: L.donovani DNA.
FEATURES Location/Qualifiers
1..3105
/organism="Leishmania donovani"
/mol_type="genomic DNA"
/db_xref="taxon:5661"
gene 101..1873
/gene="gp63"
101..1873
/gene="gp63"
/codon_start=1
/product="glycoprotein 63"
/protein_id="AAA29244.1"

sig_peptide 132..248
/gene="GP63"
mat_peptide 423..2048
/gene="GP63"
/product="surface protease"
/note="constitutive glycoprotein 63"
2621
/gene="GP63"
ORIGIN
Query Match 59.2%; Score 28.4; DB 3; Length 2621;
Best Local Similarity 76.1%; Pred. No. 6.8;
Matches 35; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
OY 1 TATGATCAGCTGTTACCGCTGTTTACCCATGAATGGCACATG 46
Db 882 TAGCACCAGCTGTGACGCTGTGTCTACGACGAGATGGCCGACG 927
RESULT 15
LTU495010 2050 bp DNA linear INV 22-JUL-2002
LOCUS Leishmania turanica mspc gene for GP63, strain
DEFINITION MRHO/SU/1983/MARZ-051.
ACCESSION AJ495010
VERSION AJ495010.1 GI:21954471
KEYWORDS GP63; major surface protease; mspc gene.
SOURCE Leishmania turanica
ORGANISM Leishmania turanica
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Leishmania.
REFERENCE 1
AUTHORS Mauricio, I.L., Stothard, J.R. and Miles, M.A.
TITLE Genetic diversity in the Leishmania donovani complex
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2050)
AUTHORS Mauricio, I.L.
TITLE Direct Submission
JOURNAL Submitted (11-JUL-2002) Mauricio I.L., Infectious and Tropical
Diseases, London School of Hygiene and Trop. Med., Keppel Street,
WC1E 7HT, UNITED KINGDOM
FEATURES
source Location/Qualifiers
1. 2050
/organism="Leishmania turanica"
/mol_type="genomic DNA"
/strain="MRHO/SU/1983/MARZ-051"
/db_xref="taxon:62297"
15..1949
/gene="mspc"
15..1949
/gene="mspc"
/note="constitutive major surface protease"
/codon_start=1
/allele="mspc.tu1"
/product="GP63"
/protein_id="CAD42819.1"
/db_xref="GI:21954472"
/db_xref="GOA:Q8KNY9"
/db_xref="SPTRMBL:Q8KNY9"
/translation="MSVDSSTHRRRCVAAIRLVLAAGVAVTVAGTAAAHAGAL
OHRGVDAHQARVQSVADHHTAPGAVSAVGPYTLDAHTAAADPRGARGVER
DVNMGALRTAVSIEDLTDPYHRCVYRGQVYNNHAGAVTCTABDILTDEKDIIVKHL
IPRALIHTERLKVQVQGTWKTGVNDEICDFKVOAHITKFSNTDFWYVASVP
SERGVLAWATTCQVFSBQHPAVGVINIPANIVSRYDQLVTRVTHMAHALGFSQPF
FEDARIIVANISVNRGNFDPVINSSTAVAKAREQCDTLEYLEVEDQAGSAGSH
IKRRNAODELMAPAAAGYATLIMAIPODLGPGOAPESKAEVLMGRNAGCALITDK
CMEONITOMPAMCNESEDAIRCPSTRILIGCGIREVETPLPIYQVPTNASLGGYS
PFLDYGPSVIGVAHGKCNODASAAELIAFVNSPAACIDGATTPKRTANGYG
GLCANVQCDTARTTYSVQVHGSDTANCTPGILVELISTVSNTEEGGYITCPYVEVC
QGNVKGAPFDGSDSSSAGDDADKAAMERNDRMAGLATAVTVLIGIVLSIMALIV
VWLLIVSCPRWCCKVGLPT"
15..131
/gene="mspc"
sig_peptide

mat_peptide 132..1832
/gene="mspc"
/product="GP63"
ORIGIN
Query Match 57.9%; Score 27.8; DB 3; Length 2050;
Best Local Similarity 74.5%; Pred. No. 11;
Matches 35; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
OY 1 TATGATCAGCTGTTACCGCTGTTTACCCATGAATGGCACATGC 47
Db 774 TAGCACCAGCTGTGACGCTGTGTCTACGACGAGATGGCCGACG 820
Search completed: March 15, 2004, 22:49:58
Job time : 210.439 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 25.7457 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 1 tatgatcagctggtaccg.....ccatgaatgcacatgca 48

Sequence: IDENTITY_NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 3373663 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_29Jan04.*

1: geneseqn1980s.*
2: geneseqn1990s.*
3: geneseqn2000s.*
4: geneseqn2001as.*
5: geneseqn2001bs.*
6: geneseqn2002s.*
7: geneseqn2003as.*
8: geneseqn2003bs.*
9: geneseqn2003cs.*
10: geneseqn2004s.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	48	3 AAC64621	AAC64621 Leishmani
2	48	100.0	78	3 AAC64609	AAC64609 Agfa (SEF
3	48	100.0	78	3 AAC64610	AAC64610 Agfa (SEF
4	48	100.0	456	3 AAC64628	AAC64628 Agfa:PT3
5	48	100.0	456	3 AAC64622	AAC64622 Agfa:PT3
6	48	100.0	456	3 AAC64626	AAC64626 Agfa:PT3
7	48	100.0	456	3 AAC64630	AAC64630 Agfa:PT3
8	48	100.0	456	3 AAC64629	AAC64629 Agfa:PT3
9	48	100.0	456	3 AAC64625	AAC64625 Agfa:PT3
10	48	100.0	456	3 AAC64631	AAC64631 Agfa:PT3
11	48	100.0	456	3 AAC64623	AAC64623 Agfa:PT3
12	48	100.0	456	3 AAC64627	AAC64627 Agfa:PT3
13	48	100.0	456	3 AAC64624	AAC64624 Agfa:PT3
14	43.2	90.0	48	3 AAC64616	AAC64616 S. enteri
15	40.6	84.6	78	3 AAC64606	AAC64606 SeFa (SEF
16	40.6	84.6	78	3 AAC64605	AAC64605 SeFa (SEF
17	27.8	57.9	2547	7 AAV52244	AAV52244 S. pneumo
18	24.6	51.3	1780	2 AAV52244	AAV52244 S. pneumo
19	24.6	51.3	2544	7 ABQ6472	ABQ6472 S. pneumo
20	24.6	51.3	2547	7 AAV52211	AAV52211 Prokaryot
21	24.6	51.3	3636	2 AAV52211	AAV52211 Streptoco
22	24.6	51.3	110000	7 ABS56454_06	Continuation (7 of
23	23	47.9	420	6 ABQ69493	ABQ69493 Listeria

C	24	23	47.9	420	6	ABQ67505	ABQ67505 Listeria
C	25	23	47.9	436	8	ACH39062	ACH39062 Human foe
C	26	23	47.9	810	5	AAH52245	AAH52245 Human AFP
C	27	23	47.9	951	1	ABL21387	ABL21387 Drosophill
C	28	23	47.9	1722	6	ABQ76709	ABQ76709 Human mac
C	29	23	47.9	2940	2	AAT36490	AAT36490 Lactococc
C	30	23	47.9	3333	4	ABL21386	ABL21386 Drosophill
C	31	23	47.9	4850	7	ABQ83879	ABQ83879 Human MDD
C	32	23	47.9	18949	4	AAK70445	AAK70445 Human immu
C	33	23	47.9	63020	6	ABQ67197_11	Continuation (12 o
C	34	23	47.9	110000	6	ABQ69245_12	Continuation (13 o
C	35	22.8	47.5	1323	3	AAF13368	AAF13368 Aspergilli
C	36	22.6	47.1	1767	7	ACA37452	ACA37452 Prokaryot
C	37	22.6	47.1	3122	2	AAQ79913	AAQ79913 Lys-amino
C	38	22.4	46.7	1014	4	AAE44754	AAE44754 Nucellin-
C	39	22.4	46.7	1014	9	ADB94941	ADB94941 PCD pathw
C	40	22.4	46.7	1925	9	ADB94853	ADB94853 PCD pathw
C	41	22.2	46.2	912	5	AAH67547	C glutami
C	42	22.2	46.2	1035	4	AAF71443	Corynebact
C	43	22.2	46.2	44147	6	ABK84481	Human CDN
C	44	22.2	46.2	44147	9	ADD14691	Human SRC
C	45	22.2	46.2	113604	8	ADA42607	Human LQT

ALIGNMENTS

RESULT 1	
ID AAC64621	standard; DNA; 48 BP.
AC AAC64621;	
DT 26-FEB-2001	(first entry)
DE Leishmania major PT3 epitope DNA sequence SEQ ID NO:9.	
XX	
KM Salmonella; agfa: chromosomal gene replacement; fimbrin; epitope;	
KW vaccine; immune response; immunogen; ds.	
XX	
OS Leishmania major.	
XX	
PN WO200060102-A2.	
XX	
PD 12-OCT-2000.	
XX	
PF 05-APR-2000; 2000MO-CA000356.	
XX	
PR 05-APR-1999; 99US-0127888P.	
XX	
PA (UTVI-) UNIV VICTORIA.	
PI White AP, Doran JL, Collison SK, Kay WW;	
XX	
XX WPI: 2000-672631/65.	
XX P-PSDB; AAB36345.	
DR	
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence	
PT which encodes foreign epitope or antigen, expresses recombinant Agfa	
PT protein useful for eliciting immune response in animal.	
XX	
PS Disclosure; Page 135; 13pp; English.	
XX	
XX The present invention describes a recombinant agfa gene (I) where a	
CC segment of the gene has been replaced by a segment of a foreign DNA	
CC sequence which encodes a foreign epitope or antigen. Also described are:	
CC (1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended	
CC assembly system of strains of Salmonella, Escherichia coli and	
CC Enterobacteriaceae for the production of fimbriae comprising recombinant	
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)	
CC directing recombination of a recombinant gene into the chromosome of the	
CC homologous species; (3) directing recombination of a recombinant gene	
CC back into the chromosome of the homologous species, replacing the native	

CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fibrillar presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fibrillae are easy and
CC the exemplification of the present invention

Sequence 48 BP; 12 A; 11 C; 11 G; 14 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 48;
Best Local Similarity 100.0%; Pred. No. 8.4e-10;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 48
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 48

RESULT 2

AAC64609/c
ID AAC64609 standard; DNA; 78 BP.

AC AAC64609;

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)

DE Agfa (SEF17) recombinant agfa::PT3a generating PCR primer 17-B.

KW Salmonella; agfa; chromosomal gene replacement; fimbtrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.

OS Salmonella enteritidis.
OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.

DR WPI; 2000-672631/65.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Example 2; Page 61, 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Agfa, Cgga and Agfa-homologue fimbtrin subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fibrillar presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell) the hybrid fibrin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fibrillar subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fibrillae are easy and
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct OS field.)

Sequence 78 BP; 17 A; 17 C; 21 G; 23 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;
Best Local Similarity 100.0%; Pred. No. 9.5e-10;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 48
48 TATGATCAGCTGTTACCCGTTGTTACCATGAATGCGACATGCA 1

RESULT 3

AAC64610
ID AAC64610 standard; DNA; 78 BP.

AC AAC64610;

DT 06-AUG-2003 (revised)
DT 26-FEB-2001 (first entry)

DE Agfa (SEF17) recombinant agfa::PT3a generating PCR primer 17-C.

KW Salmonella; agfa; chromosomal gene replacement; fimbtrin; epitope;
KW vaccine; immune response; immunogen; PCR primer; ss.

OS Salmonella enteritidis.
OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
PI WPI; 2000-672631/65.

DR WPI; 2000-672631/65.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX Example 2; Page 61, 139p; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
CC correct US field.)

XX Sequence 78 BP; 21 A; 19 C; 17 G; 21 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 78;

Best Local Similarity 100.0%; Pred. No. 9.5e-10; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGACATGCA 48
Db 1 TATGATCAGCTGTTTACCCGTTGTTTACCCATGAATGACATGCA 48

RESULT 4
AAC64628

ID AAC64628 standard; DNA; 456 BP.

XX AAC64628;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#7 DNA sequence SEQ ID NO:23.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen; ds.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX P-PSDB; AAB36352.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF1/TA) nucleation depended
CC assembly; system of strains of *Salmonella*, *Escherichia coli* and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 100.0%; Score 48; DB 3; Length 456;

Best Local Similarity 100.0%; Pred. No. 1.5e-09; Indels 0; Gaps 0;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGACATGCA 48
Db 331 TATGATCAGCTGTTACCCGTTGTTTACCCATGAATGACATGCA 378

RESULT 5
AAC64622

ID AAC64622 standard; DNA; 456 BP.

XX AAC64622;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#1 DNA sequence SEQ ID NO:11.

XX Salmonella; agfa; chromosomal gene replacement; fimbrial; epitope;

KW vaccine; immune response; immunogen; ds.

XX Salmonella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO2000060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

XX P-PSDB; AAB36346.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 135, 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGGTACCCGTTGTGTTACCCATGAATGCGACATGCA 48
DB 382 TATGATCAGCTGGTACCCGTTGTGTTACCCATGAATGCGACATGCA 429

RESULT 6
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX

DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 DNA sequence SEQ ID NO:19.

XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen; ds.
OS *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.

XX WO20060102-A2.
XX
PD 12-OCT-2000.

XX
PP 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.
XX
PA

XX White AP, Doran JL, Collison SK, Kay MW;
XX MPI; 2000-672631/65.
XX P-PSDB; AAB36350.
XX
PI
DR
DR
DR
DR
PT

Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137, 139pp; English.

XX The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGGTACCCGTTGTGTTACCCATGAATGCGACATGCA 48
DB 196 TATGATCAGCTGGTACCCGTTGTGTTACCCATGAATGCGACATGCA 243

RESULT 7
AAC64630
ID AAC64630 standard; DNA; 456 BP.
XX
AC AAC64630;
XX

DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#9 DNA sequence SEQ ID NO:27.

XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen; ds.
OS *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.

XX WO20060102-A2.
XX
PD 12-OCT-2000.

XX
PP 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.

XX (UYVI-) UNIV VICTORIA.
XX
PA

XX White AP, Doran JL, Collison SK, Kay MW;
XX MPI; 2000-672631/65.
XX
PI
DR
DR
DR
DR
PT

DR P-PSDB; AAB36354.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CagA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGCAATGCA 48
Db 241 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGCAATGCA 288
XX
RESULT 8
AAC64629
ID AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AGFA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX

PI White AP, Doran JL, Collison SK, Kay MW;
XX
XX WPI; 2000-672631/65.
DR
DR P-PSDB; AAB36353.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CagA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
XX
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 1 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGCAATGCA 48
Db 172 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGCAATGCA 219
XX
RESULT 9
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE AGFA::PT3#4 DNA sequence SEQ ID NO:17.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX *Salmonella enteritidis*.
OS *Escherichia coli*.
OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX

XX (UYVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WM,
 PI MPI, 2000-672631/65.
 XX P-PSDB; AAB36349.
 DR
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGTATACCCGTTGTTTACCATGAATGGACATGCA 48
 DB 127 TATGATCAGCTGTATACCCGTTGTTTACCATGAATGGACATGCA 174
 RESULT 10
 AAC64631
 ID AAC64631 standard; DNA; 456 BP.
 XX
 AC AAC64631;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX

PF 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 PR
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX
 PI White AP, Doran JL, Collison SK, Kay WM,
 PT MPI, 2000-672631/65.
 DR P-PSDB; AAB36355.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGTATACCCGTTGTTTACCATGAATGGACATGCA 48
 DB 307 TATGATCAGCTGTATACCCGTTGTTTACCATGAATGGACATGCA 354
 RESULT 11
 AAC64623
 ID AAC64623 standard; DNA; 456 BP.
 XX
 AC AAC64623;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#2 DNA sequence SEQ ID NO:13.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 PN WO200060102-A2.
 XX

XX 12-OCT-2000.
PD 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
DR WPI; 2000-672631/65.
P-PSDB; AAB36347.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombinant of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombinant of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGACATGCA 48
Db 352 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGACATGCA 359
RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
XX
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE AGFA::PT3#6 DNA sequence SEQ ID NO:21.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS *Salmonella enteritidis*.
OS *Escherichia coli*.

OS Synthetic.
XX
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
PD 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
DR WPI; 2000-672631/65.
P-PSDB; AAB36351.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF17/TAF) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombinant of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombinant of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;
Query Match 100.0%; Score 48; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.5e-09;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGACATGCA 48
Db 262 TATGATCAGCTGTTACCGTGTGTTACCCATGAATGACATGCA 309
RESULT 13
AAC64624
ID AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AGFA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KM *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS

XX OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 XX Synthetic.
 XX MO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 DR WPI; 2000-672631/65.
 XX P-PSDB; AAB36348.
 XX
 PS Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SO Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 Query Match 100.0%; Score 48; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 1.5e-09;
 Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 48
 DB 67 TATGATCAGCTGCTTACCGTGTGTGTACCATGAATGCGACATGCA 114
 RESULT 14
 AAC64616
 ID AAC64616 standard; DNA; 48 BP.
 XX AAC64616;
 AC
 XX
 DT 06-AUG-2003 (revised)
 XX 26-FEB-2001 (first entry)
 XX

DE S. enteritidis recombinant agfa gene sequence.
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 XX OS *Salmonella enteritidis*.
 OS Synthetic.
 XX MO200060102-A2.
 XX 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 XX White AP, Doran JL, Collison SK, Kay MW;
 DR WPI; 2000-672631/65.
 XX
 PS Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Fig 1c; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous species; (3) directing recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention. (Updated on 06-AUG-2003 to
 CC correct OS field.)
 XX
 SO Sequence 48 BP; 14 A; 10 C; 10 G; 14 T; 0 U; 0 Other;
 Query Match 90.0%; Score 43.2; DB 3; Length 48;
 Best Local Similarity 93.8%; Pred. No. 7.6e-08;
 Matches 45; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 TATGATCAGCTGCTTACCGTGTGTACCATGAATGCGACATGCA 48
 DB 1 TATGATCAGCTGCTTACCGTGTGTGTACCATGAATGCGACATGCA 48
 RESULT 15
 AAC64606
 ID AAC64606 standard; DNA; 78 BP.
 XX AAC64606;
 AC
 XX
 DT
 XX

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 5.01734 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcagctggttacccg.....cccatgaatgacacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 27747546 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUTS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	24.6	51.3	3636	US-08-961-527-78	Sequence 78, Appl
C 2	22.4	46.7	1014	US-09-325-932A-15	Sequence 15, Appl
C 3	22	45.8	5406	US-08-961-527-166	Sequence 166, Appl
C 4	21.8	45.4	7754	US-09-614-238-3	Sequence 3, Appl
C 5	21.4	44.6	648	US-09-252-991A-11586	Sequence 11586, A
C 6	21.4	44.6	963	US-09-252-991A-11645	Sequence 11645, A
C 7	21.4	44.6	987	US-09-252-991A-11794	Sequence 11794, A
C 8	21.4	44.6	1983	US-09-453-702B-36	Sequence 36, Appl
C 9	21.4	44.6	2166	US-08-317-401E-3	Sequence 3, Appl
C 10	21.4	44.6	2877	US-08-317-401E-1	Sequence 1, Appl
C 11	21.4	44.6	9541	US-08-961-527-132	Sequence 132, Appl
C 12	21	43.8	2975	US-09-543-681A-3167	Sequence 3167, Appl
C 13	21	43.8	6587	US-08-879-337-3	Sequence 3, Appl
C 14	21	43.8	100848	US-09-596-002-39	Sequence 39, Appl
C 15	21	43.8	705	US-09-543-681A-1102	Sequence 1102, Appl
C 16	20.8	43.3	1029	US-09-328-352-2205	Sequence 2205, Appl
C 17	20.8	42.9	1239	US-09-328-352-779	Sequence 779, Appl
C 18	20.6	42.9	1872	US-09-345-650-2	Sequence 2, Appl
C 19	20.6	42.9	3366	US-09-543-681A-1952	Sequence 1952, Appl
C 20	20.6	42.9	8334	US-09-621-976-8341	Sequence 8341, Appl
C 21	20.6	42.5	440	US-09-252-991A-13564	Sequence 13564, A
C 22	20.4	42.5	753	US-09-252-991A-13860	Sequence 13860, A
C 23	20.4	42.5	1404	US-09-107-532A-2521	Sequence 2521, Appl
C 24	20.4	42.5	1608	US-09-328-352-2416	Sequence 2416, Appl
C 25	20.4	42.5	1848	US-08-637-670-23	Sequence 23, Appl

C 28	20.2	42.1	1203	US-09-107-532A-2612	Sequence 2612, Appl
C 29	20	41.7	562	US-09-328-111-468	Sequence 468, Appl
C 30	20	41.7	1152	US-09-489-039A-2755	Sequence 2755, Appl
C 31	20	41.7	1280	US-08-335-844A-10	Sequence 10, Appl
C 32	20	41.7	1280	US-09-129-366-10	Sequence 10, Appl
C 33	20	41.7	1293	US-08-335-844A-11	Sequence 11, Appl
C 34	20	41.7	1293	US-09-129-366-11	Sequence 11, Appl
C 35	20	41.7	2472	US-08-335-844A-7	Sequence 7, Appl
C 36	20	41.7	2472	US-09-129-366-7	Sequence 7, Appl
C 37	20	41.7	3213	US-09-107-532A-62	Sequence 62, Appl
C 38	20	41.7	3296	US-08-335-844A-14	Sequence 14, Appl
C 39	20	41.7	3296	US-09-129-366-14	Sequence 14, Appl
C 40	20	41.7	3305	US-08-335-844A-8	Sequence 8, Appl
C 41	20	41.7	3305	US-09-129-366-8	Sequence 8, Appl
C 42	20	41.7	3319	US-08-335-844A-15	Sequence 15, Appl
C 43	20	41.7	3319	US-09-129-366-15	Sequence 15, Appl
C 44	20	41.7	3358	US-08-335-844A-20	Sequence 20, Appl
C 45	20	41.7	3358	US-09-129-366-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1
US-08-961-527-78/C
Sequence 78, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunsch

TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences

NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 3636 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-78

Query Match

Best Local Similarity 76.9%; Pred. No. 1.1;

Matches 30; Conservative 0; Mismatches 9; Indels 0;

6 TCAAGTGTACCGCTGTGTACCAATGCAATGCA 44

Db 2806 TCAACAGTGTACCGCTGTGTGTGCGCCATGATGCTCA 2768

RESULT 2

US-09-325-932A-15
; Sequence 15, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; FILE REFERENCE: 1022
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Pinus radiata
US-09-325-932A-15

Query Match

Best Local Similarity 46.7%; Score 22.4; DB 4; Length 1014;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 1 TATGATGAGCTGCTTACCCGCTGTTGTTACCCATGAATGCGACATGCA 48

DB 439 TATGACGACGAGGCTTCTTTCAGTTTACCTGCATGACATGCA 486

RESULT 3

US-08-961-527-166/c
; Sequence 166, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae polynucleotides and sequences
; CORRESPONDENCE ADDRESS: 391
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 166:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5406 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-166

Query Match

45.8%; Score 22; DB 4; Length 5406;

Best Local Similarity 67.4%; Pred. No. 15;
Matches 31; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 2 ATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGCGACATGC 47

DB 3442 AGATCAGCTAAGACGCTTATGATTAATCATCATGAAATGCG 3397

RESULT 4

US-09-634-238-3/c
; Sequence 3, Application US/09634238
; Patent No. 6544772
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka J.
; APPLICANT: Bloksberg, Leonard, N.
; APPLICANT: Lubbers, Mark W.
; APPLICANT: Dekker, James
; APPLICANT: Christensen, Anna C.
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul W.
; APPLICANT: Reid, Julian R.
; TITLE OF INVENTION: Polynucleotides, materials incorporating
; FILE REFERENCE: 11000,1043U1
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 422
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 7754
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-634-238-3

Query Match

Best Local Similarity 45.4%; Score 21.8; DB 4; Length 7754;
Matches 29; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 5 ATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGCGACAT 45

DB 4023 AGCAGCGGCTTCTTACCGCTTATGCGCATGAACTGCCCAT 3983

RESULT 5

US-09-252-991A-11586/c
; Sequence 11586, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196,136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 3142
; SEQ ID NO 11586
; LENGTH: 648
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11586

Query Match

Best Local Similarity 44.6%; Score 21.4; DB 4; Length 648;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 4 GATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGCGA 42

Db 296 GATCGCTCGATCTGTGTTTCCCATGACTGAA 258

RESULT 6

US-09-252-991A-11645/C
; Sequence 11645, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11645
; LENGTH: 963
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11645

Query Match 44.6%; Score 21.4; DB 4; Length 963;

Best Local Similarity 71.8%; Pred. No. 16; Mismatches 11; Indels 0; Gaps 0;

Db 316 GATCGCTCGATCTGTGTTTCCCATGACTGAA 278

Query 4 GATCGCTGTTACCCGTTGTTTACCAGTAATGCA 42

US-09-252-991A-11794
; Sequence 11794, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11794
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11794

Query Match 44.6%; Score 21.4; DB 4; Length 987;

Best Local Similarity 71.8%; Pred. No. 16; Mismatches 11; Indels 0; Gaps 0;

Db 681 GATCGCTCGATCTGTGTTTCCCATGACTGAA 719

RESULT 8

US-09-453-702B-36/C
; Sequence 36, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blatterer, Frederick R.
; Burland, Valerie
; Perna, Nicole T.

plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady

STREET: 1 South Pinckney Street

CITY: Madison

STATE: WI

COUNTRY: US

ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960236.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 36:

SEQUENCE CHARACTERISTICS:

LENGTH: 1983

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 36:

US-09-453-702B-36

Query Match 44.6%; Score 21.4; DB 4; Length 1983;

Best Local Similarity 66.0%; Pred. No. 20; Mismatches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Db 372 ATGATCACTGTTACCCGTTGTTTACCAGTAATGCAATGCA 48

Query 2 ATGATCACTGTTACCCGTTGTTTACCAGTAATGCAATGCA 48

US-08-317-401E-3
; Sequence 3, Application US/08317401E
; Patent No. 5922561
; GENERAL INFORMATION:
; APPLICANT: Thompson, Sheryl Ann

TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5922561o No. 5922561th America, Inc.

STREET: 405 Lexington Avenue, Suite 6400

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25 (BEO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/317,401E

Thu Mar 18 12:28:09 2004

us-09-543-407-9.rn1

Page 4

FILING DATE: 03-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4248.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2166 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
US-08-317-401E-3

Query Match
Best Local Similarity 44.6%; Score 21.4; DB 2; Length 2166;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 4 GATGAGCTGTTACCCGTTGTTGTTACCCATG 34
Db 326 GATGAGCTGTTACCCGTTGTTGTTACCCATG 356

RESULT 10
US-08-317-401E-1
Sequence 1, Application US/08317401E
Patent No. 5922561
GENERAL INFORMATION:
APPLICANT: Thompson, Sheryl Ann
APPLICANT: Javer, Debbie Sue
TITLE OF INVENTION: GENES ENCODING SIGNAL RECOGNITION PARTICLE OF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESSES:
ADDRESSEE: No. 59225610 No. 5922561disk of No. 5922561th America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08317,401E
FILING DATE: 03-October-1994
ATTORNEY/AGENT INFORMATION:
NAME: Harrington, James J.
REGISTRATION NUMBER: 38,711
REFERENCE/DOCKET NUMBER: 4248.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 867 0123
TELEFAX: 212 867 0298
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2877 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Aspergillus niger
FEATURE:
NAME/KEY: CDS
LOCATION: join (126..203, 253..1776)
US-08-317-401E-1

Query Match
Best Local Similarity 44.6%; Score 21.4; DB 2; Length 2877;
Matches 25; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
Qy 4 GATGAGCTGTTACCCGTTGTTGTTACCCATG 34
Db 418 GATGAGCTGTTACCCGTTGTTGTTACCCATG 448

RESULT 11
US-08-961-527-132/c
Sequence 132, Application US/08961527
Patent No. 6420135
GENERAL INFORMATION:
APPLICANT: Charles Kunsch
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 132:
SEQUENCE CHARACTERISTICS:
LENGTH: 9541 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-08-961-527-132

Query Match
Best Local Similarity 44.6%; Score 21.4; DB 4; Length 9541;
Matches 28; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
Qy 7 CAGTGTATGACCGTGTGTTGTTACCGATGATGCAAT 45
Db 6099 CAGTGTATGACCGTGTGTTGTTACCGATGATGCAAT 6061

RESULT 12
US-09-543-681A-3167
Sequence 3167, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706

;; PRIOR FILING DATE: 1999-04-09
;; NUMBER OF SEQ ID NOS: 8344
;; SEQ ID NO 3167
;; LENGTH: 597
;; TYPE: DNA
;; ORGANISM: Proteus mirabilis
US-09-543-681A-3167

Query Match 43.8%; Score 21; DB 4; Length 597;
Best Local Similarity 73.0%; Pred. No. 21;
Matches 27; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 12 GGTACCCGCTGTTTACCCATGAATGCACATCA 48
Db 375 GGTGAACTGTGCTCACTTGAATGACACAGCA 411

RESULT 13
US-08-879-337-10
; Sequence 10, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
; APPLICANT: Jang, Yvan-Chyun
; TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
; FILE REFERENCE: 00786/338001
; CURRENT APPLICATION NUMBER: US/08/879,337B
; CURRENT FILING DATE: 1997-06-20
; EARLIER APPLICATION NUMBER: 60/022,086
; EARLIER FILING DATE: 1996-06-21.
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 2975
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(2975)
; OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-10

Query Match 43.8%; Score 21; DB 4; Length 2975;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGCTGTTTACCCATGAATGCACAT 45
Db 310 TTTATCAGATGGCTCTCACTTGTTCATCGATGCAATGTATAT 354

RESULT 14
US-08-879-337-3
; Sequence 3, Application US/08879337B
; Patent No. 6639130
; GENERAL INFORMATION:
; APPLICANT: Jang, Yvan-Chyun
; APPLICANT: Sheen, Jen
; TITLE OF INVENTION: PLANT STEROL REDUCTASES AND USES THEREOF
; FILE REFERENCE: 00786/338001
; CURRENT APPLICATION NUMBER: US/08/879,337B
; CURRENT FILING DATE: 1997-06-20
; EARLIER APPLICATION NUMBER: 60/022,086
; EARLIER FILING DATE: 1996-06-21
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 6587
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: variation
; LOCATION: (1)...(6587)

;; OTHER INFORMATION: N can be any nucleotide.
US-08-879-337-3

Query Match 43.8%; Score 21; DB 4; Length 6587;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGTTACCCGCTGTTTACCCATGAATGCACAT 45
Db 3664 TTTATCAGATGGCTCTCACTTGTTCATCGATGCAATGTATAT 3708

RESULT 15
US-09-596-002-39
; Sequence 39, Application US/09596002
; Patent No. 6632636
; GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; CURRENT FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; PRIOR FILING DATE: 1999-06-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PERL Program
; SEQ ID NO 39
; LENGTH: 100848
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte template ID No. 6632636 39
; PUBLIC INFORMATION:
US-09-596-002-39

Query Match 43.8%; Score 21; DB 4; Length 100848;
Best Local Similarity 66.7%; Pred. No. 86;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 4 GATCAGCTGTTACCCGCTGTTTACCCATGAATGCACATGCA 48
Db 73642 GATGTGCTATTGACACATATTTATCATCAATGCAACAGA 73686

Search completed: March 16, 2004, 04:37:03
Job time : 7.01734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 42.2775 seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48
Sequence: 1 taatgacgcgtgtaccgcg.....ccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCRT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCRTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27.8	57.9	2547	12 US-10-282-122A-36154	Sequence 36154, A
2	24.6	51.3	2547	12 US-10-282-122A-37710	Sequence 37710, A
3	24.6	51.3	3636	12 US-10-158-844-78	Sequence 78, App1
4	24	50.0	409	12 US-10-425-599-119735	Sequence 119735, A
5	23.2	48.3	720	12 US-10-425-114-6010	Sequence 6010, App
6	23	47.9	420	15 US-10-398-221-318	Sequence 318, App
7	23	47.9	420	15 US-10-398-221-3306	Sequence 2306, App
8	23	47.9	436	10 US-09-918-995-26274	Sequence 26274, A
9	23	47.9	1163020	15 US-10-398-221-1058	Sequence 10, App1
10	23	47.9	3011208	15 US-10-282-122A-25322	Sequence 25322, App
11	22.6	47.1	1767	12 US-10-282-122A-25322	Sequence 25322, A
12	22.6	47.1	3984	15 US-10-369-493-42554	Sequence 42554, A
13	22.4	46.7	1014	14 US-10-219-220-15	Sequence 15, App1
14	22.4	46.7	1925	14 US-10-219-220-217	Sequence 217, App
15	22.2	46.2	912	9 US-09-738-626-2582	Sequence 2582, App

c 16	22.2	46.2	2370	15 US-10-369-493-32121	Sequence 32121, A
c 17	22.2	46.2	2487	15 US-10-369-493-44598	Sequence 44598, A
c 18	22.2	46.2	2574	15 US-10-369-493-34285	Sequence 34285, A
c 19	22.2	46.2	113604	14 US-10-227-195A-1	Sequence 1, App1
c 20	22.2	46.2	113604	14 US-10-227-195A-2	Sequence 2, App1
c 21	22.2	46.2	3309400	9 US-09-738-626-1	Sequence 1, App1
c 22	22.2	45.8	145	9 US-09-864-761-24629	Sequence 24629, A
c 23	22.2	45.8	480	14 US-10-326-956-320	Sequence 320, App
c 24	22.2	45.8	557	9 US-09-864-761-7925	Sequence 7925, App
c 25	22.2	45.8	1791	9 US-09-822-830A-241	Sequence 241, App
c 26	22.2	45.8	5406	12 US-10-158-844-166	Sequence 166, App
c 27	21.8	45.4	1470	12 US-10-424-599-56999	Sequence 56999, A
c 28	21.8	45.4	1510	12 US-10-425-114-36154	Sequence 36154, A
c 29	21.8	45.4	1646	12 US-10-425-114-31422	Sequence 31422, A
c 30	21.8	45.4	2394	12 US-10-425-114-32979	Sequence 32979, A
c 31	21.8	45.4	3328	12 US-10-425-114-34623	Sequence 34623, A
c 32	21.8	45.4	7754	15 US-10-264-213-3	Sequence 3, App1
c 33	21.8	45.4	8495	13 US-10-108-605-356	Sequence 356, App
c 34	21.6	45.0	633	9 US-09-917-800A-1050	Sequence 1050, App
c 35	21.6	45.0	729	9 US-09-738-626-1352	Sequence 1352, App
c 36	21.6	45.0	852	12 US-10-627-476-367	Sequence 367, App
c 37	21.6	45.0	1580	12 US-10-425-114-21821	Sequence 21821, A
c 38	21.6	45.0	1746	15 US-10-260-238-578	Sequence 578, App
c 39	21.6	45.0	2853	12 US-10-425-114-13682	Sequence 13682, A
c 40	21.6	45.0	1691139	14 US-10-067-514-1	Sequence 1, App1
c 41	21.6	45.0	1591139	15 US-10-419-723-1	Sequence 1, App1
c 42	21.6	45.0	3309400	9 US-09-738-626-1	Sequence 1, App1
c 43	21.4	44.6	1011	12 US-10-282-122A-30293	Sequence 30293, A
c 44	21.4	44.6	1179	10 US-09-769-744A-119	Sequence 119, App
c 45	21.4	44.6	1983	14 US-10-114-170-36	Sequence 36, App1

ALIGNMENTS

RESULT 1

US-10-282-122A-36154

Sequence 36154, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Karl

APPLICANT: Zykkind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: EITRA.034A

CURRENT APPLICATION NUMBER: US/10/282, 122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191, 078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206, 848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207, 727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230, 335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230, 347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242, 578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253, 625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257, 931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267, 636

;; PRIOR FILING DATE: 2001-02-09
;; PRIOR APPLICATION NUMBER: 60/269,308
;; PRIOR FILING DATE: 2001-02-16
;; Remaining Prior Application data removed - See File Wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 78614
;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO: 36154
;; LENGTH: 2547
;; TYPE: DNA
;; ORGANISM: Streptococcus mutans
US-10-282-122A-36154

Query Match 57.9%; Score 27.8; DB 12; Length 2547;
Best Local Similarity 82.1%; Pred. No. 0.27;
Matches 32; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 7 CAGCTGTTACCGGTGTTTCCCATGAATGGCAAT 45
DB 850 CAGCAGTTGCACCTGTGTGGCCCATGACTGGCGCAT 888

RESULT 2
US-10-282-122A-37710
; Sequence 37710, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 37710
; LENGTH: 2547
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-10-282-122A-37710

Query Match 51.3%; Score 24.6; DB 12; Length 2547;
Best Local Similarity 76.9%; Pred. No. 6;

Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 6 TCAGCTGTTACCGGTGTTTCCCATGAATGGCACA 44
DB 849 TCACAGTTGCCCTTGTGTGGCCCATGAATGGCTCA 887

RESULT 3
US-10-158-844-78/c
; Sequence 78, Application US/10158844
; Publication No. US20040029118A1
; GENERAL INFORMATION:
; APPLICANT: Kunsch et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-R
; COMPUTER: Dell Latitude Pentium 3
; OPERATING SYSTEM: Windows 98
; SOFTWARE: ASCII text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/158,844
; FILING DATE: 03-Jun-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/961,527
; FILING DATE: 1997-10-30
; APPLICATION NUMBER: US 60/029,960
; FILING DATE: 1996-10-31
; ATTORNEY/AGENT INFORMATION:
; NAME: Hyman, Mark J.
; REGISTRATION NUMBER: 46,789
; REFERENCE/DOCKET NUMBER: PB340P1D1
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3636 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 78:
US-10-158-844-78

Query Match 51.3%; Score 24.6; DB 12; Length 3636;
Best Local Similarity 76.9%; Pred. No. 6.6;
Matches 30; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 6 TCAGCTGTTACCGGTGTTTCCCATGAATGGCACA 44
DB 2806 TCACAGTTGCCCTTGTGTGGCCCATGAATGGCTCA 2768

RESULT 4
US-10-424-599-119735/c
; Sequence 119735, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684

SEQ ID NO 119735
LENGTH: 409
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT3847_79129C.1
US-10-424-599-119735

Query Match
Best Local Similarity 50.0%; Score 24; DB 12; Length 409;
Best Local Similarity 68.8%; Pred. No. 6.2;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGACATGCA 48
Db 340 TACGGTGCACGCTTGCCCGCTTCTCGCCGAGAAAGGCCATCCA 293

RESULT 5
US-10-425-114-6010
Sequence 6010, Application US/10425114
Publication No. US2004003488A1
GENERAL INFORMATION:
APPLICANT: Liu, Jindong
APPLICANT: Zhou, Yihua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E
APPLICANT: Tabaska, Jack E
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53313)B
CURRENT APPLICATION NUMBER: US/10/425,114
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 6010
LENGTH: 720
TYPE: DNA
ORGANISM: Zea mays
FEATURE:
OTHER INFORMATION: Clone ID: 700550972_FLI
US-10-425-114-6010

Query Match
Best Local Similarity 48.3%; Score 23.2; DB 12; Length 720;
Best Local Similarity 70.5%; Pred. No. 16;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGACATGCA 45
Db 285 ATGTCAGCTGCTTCCGATCTTTTCCCATGAGAAAGGCCAT 328

RESULT 6
US-10-398-221-318
Sequence 318, Application US/10398221
Publication No. US2004001851A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 318
LENGTH: 420
TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-318

Query Match
Best Local Similarity 47.9%; Score 23; DB 15; Length 420;
Best Local Similarity 68.1%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGACATGCG 47
Db 175 TCTGATTATCAGCTTAAGTATGTACTGCTCATGAACTAGGCGATGC 221

RESULT 7
US-10-398-221-2306
Sequence 2306, Application US/10398221
Publication No. US2004001851A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
APPLICANT: GLASER, Philippe
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT APPLICATION NUMBER: US/10/398,221
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 2306
LENGTH: 420
TYPE: DNA
ORGANISM: Listeria innocua
US-10-398-221-2306

Query Match
Best Local Similarity 47.9%; Score 23; DB 15; Length 420;
Best Local Similarity 68.1%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGACATGCG 47
Db 175 TCTGATTATCAGCTTAAGTATGTACTGCTCATGAACTAGGCGATGC 221

RESULT 8
US-09-918-995-26274/C
Sequence 26274, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 26274
LENGTH: 436
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: (1)..(436)
OTHER INFORMATION: n = A,T,C or G
US-09-918-995-26274

Query Match
Best Local Similarity 47.9%; Score 23; DB 10; Length 436;
Best Local Similarity 68.1%; Pred. No. 16;
Matches 32; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGACATGCA 48
||||| ||||||| ||| ||| ||||| ||| |||

;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374
;; SEQ ID NO 42554
;; LENGTH: 3984
;; TYPE: DNA
;; ORGANISM: Anabaena PCC7120
US-10-369-493-42554

Query Match 47.1%; Score 22.6; DB 15; Length 3984;
Best Local Similarity 75.7%; Pred. No. 47;
Matches 28; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 6 TCAGCTGTTACCCGTTGTTACCCATGAATGCA 42
Db 843 TCACCTACTTACCGGTGATGATGCCATTATGTGCA 879

RESULT 13

US-10-219-220-15
; Sequence 15, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:

;; APPLICANT: Flinn, Barry
;; APPLICANT: Laasham, Annette
;; TITLE OF INVENTION: Compositions affecting programmed cell
;; TITLE OF INVENTION: death and their use in the modification of plant development
;; FILE REFERENCE: 11000.1022c1
;; CURRENT APPLICATION NUMBER: US/10/219,220
;; PRIOR FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;; PRIOR FILING DATE: 1999-06-04
;; NUMBER OF SEQ ID NOS: 290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 15
;; LENGTH: 1014
;; TYPE: DNA
;; ORGANISM: Pinus radiata
US-10-219-220-15

Query Match 46.7%; Score 22.4; DB 14; Length 1014;
Best Local Similarity 66.7%; Pred. No. 38;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 TATGATCAGTGTACCCGTTGTTACCCATGAATGCAATGCA 48
Db 439 TATGACGACGAGGTTCTTTCAGATTTCACCTGACGACGATGGA 486

RESULT 14

US-10-219-220-217
; Sequence 217, Application US/10219220
; Publication No. US20030082724A1
; GENERAL INFORMATION:

;; APPLICANT: Flinn, Barry
;; APPLICANT: Laasham, Annette
;; TITLE OF INVENTION: Compositions affecting programmed cell
;; TITLE OF INVENTION: death and their use in the modification of plant development
;; FILE REFERENCE: 11000.1022c1
;; CURRENT APPLICATION NUMBER: US/10/219,220
;; PRIOR FILING DATE: 2002-08-14
;; PRIOR APPLICATION NUMBER: U.S. No. US20030082724A1 09/325,932
;; PRIOR FILING DATE: 1999-06-04
;; NUMBER OF SEQ ID NOS: 290
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 217
;; LENGTH: 1925
;; TYPE: DNA
;; ORGANISM: Pinus radiata
US-10-219-220-217

Query Match 46.7%; Score 22.4; DB 14; Length 1925;
Best Local Similarity 66.7%; Pred. No. 46;
Matches 32; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

Qy 1 TATGATCAGTGTACCCGTTGTTACCCATGAATGCAATGCA 48
Db 980 TATGACGACGAGGTTCTTTCAGATTTCACCTGACGACGATGGA 1027

RESULT 15

US-09-738-626-2582
; Sequence 2582, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

;; APPLICANT: NAKAGAWA, SATOSHI
;; APPLICANT: MIZOGUCHI, HIROSHI
;; APPLICANT: ANDO, SEIKO
;; APPLICANT: HAYASHI, MIKIRO
;; APPLICANT: OCHIAI, KEIKO
;; APPLICANT: YOKOI, HARUHIKO
;; APPLICANT: TATEISHI, NAOKO
;; APPLICANT: SENOH, AKIHIRO
;; APPLICANT: IKEDA, MASATO
;; APPLICANT: OZAKI, AKIO
;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; PRIOR FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 2582
;; LENGTH: 912
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2582

Query Match 46.2%; Score 22.2; DB 9; Length 912;
Best Local Similarity 69.8%; Pred. No. 44;
Matches 30; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3 TGATCAGTGTATACCCGTTGTTACCCATGAATGCAATGCAAT 45
Db 180 TGATGCGCTGTCAAGCGTGTGTCATGAGAGGACAGT 222

Search completed: March 17, 2004, 08:15:38
Job time : 54.3775 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 234.832 Seconds

(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-9

Sequence: 1 tatgatcagcgcgttaccgc.....cccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:
1: em_estha:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hc1:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc3:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrc:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	25.2	52.5	878	28	CC001571 PUDIE23TD
2	25	52.1	894	13	BUI47350 AGENCOURT
3	24.8	51.7	273	10	BB435111 BB435111
4	24.6	51.3	1201	9	AL576676 AL576676

5	24.2	50.4	445	9	AU299906
6	24.2	50.4	539	28	AZ056038
7	24.2	50.4	903	14	CB200568
8	24.2	50.4	1032	29	CNS01XVV
9	24	50.0	176	10	AM837303
10	24	50.0	300	10	BF634308
11	24	50.0	600	12	BF102563
12	24	50.0	610	9	AL589140
13	24	50.0	766	13	B0294276
14	24	50.0	780	13	B0252275
15	24	50.0	791	13	B0344056
16	24	50.0	812	13	B0328355
17	24	50.0	813	13	B0202575
18	23.8	49.6	294	9	AA224363
19	23.6	49.2	296	10	AW373707
20	23.6	49.2	316	14	CB017485
21	23.6	49.2	994	13	B0250654
22	23.4	48.8	446	10	BE443880
23	23.4	48.8	605	14	CB891544
24	23.4	48.8	673	14	CD452770
25	23.4	48.8	1016	29	CC600714
26	23.2	48.3	153	13	BQ172708
27	23.2	48.3	230	10	BB435292
28	23.2	48.3	248	14	CF245156
29	23.2	48.3	275	10	BB435052
30	23.2	48.3	295	10	BB436405
31	23.2	48.3	305	10	BB433896
32	23.2	48.3	460	12	BM379062
33	23.2	48.3	467	13	BQ744659
34	23.2	48.3	479	12	BQ947566
35	23.2	48.3	521	14	CD204019
36	23.2	48.3	522	12	BI076348
37	23.2	48.3	540	9	A1711846
38	23.2	48.3	545	10	BE358333
39	23.2	48.3	577	14	CA404413
40	23.2	48.3	590	12	BG103593
41	23.2	48.3	591	12	BM318096
42	23.2	48.3	615	10	BE355335
43	23.2	48.3	620	14	CF049562
44	23.2	48.3	623	10	BE919209
45	23.2	48.3	623	14	CD527595

ALIGNMENTS

RESULT 1
CC001571
LOCUS
DEFINITION
PUDIE23TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZMMB217C21.
ACCESSION
CC001571
VERSION
CC001571.1 GI:29380131
KEYWORDS
GSS.
SOURCE
Zea_mays
ORGANISM
Zea_mays
REFERENCE
Whitehead, C.A., Quackenbush, J., Van Aken, S., Uterback, T., Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and Benneken, J.
Maize Genomics Consortium
Unpublished (2003)
CONTACT
Contact: Cathy Whitelaw
TITLE
JOURNAL
COMMENT
7912 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.

FEATURES

Location/Qualifiers

1.878
/organism="Zea mays"
/mol_type="Genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZMMBTA217C21"
/clone_1ib="ZM 0.61.0 KB"
/note="Vector: pCR4-TOPO, Site 1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN

Query Match
Best Local Similarity 52.5%; Score 25.2; DB 28; Length 878;
Matches 33; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

2 ATGATCAGCTGTTACCCGTTGTTACCAATGAAATGCAATGC 47
548 ATGACCGACAGTTGTCGACTGAAACCATGATGATGATGC 593

RESULT 2

LOCUS BUI47350

DEFINITION AGENCOURT 8070809 NIH_MGC_112 Homo sapiens CDNA clone IMAGE:6089930

5', mRNA sequence.

VERSION BUI47350.1 GI:22660882

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 894)
NIH-MGC http://mgs.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)

Contact: Robert Strauberg, Ph.D.
Email: Ggaps-remail.nih.gov
Tissue Procurement: DCTD/DRP

CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNLN)

Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov

Plate: LNCM2330 row: 1 column: 03
High quality sequence stop: 723.

Location/Qualifiers

1.894
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6089930"
/tissue_type="melanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1ib="NIH MGC 113"
/note="Organ: skin; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC library."

FEATURES

source

ORIGIN

Query Match
Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

Db

841 ATGACGACCCCTTACCACTGTTTACCAACCAAGGCGCA 881

RESULT 3

LOCUS BB435111

DEFINITION

BB435111 RIKEN full-length enriched, adult pancreas islet cells Mus
musculus CDNA clone C820011N02 3', mRNA sequence.

ACCESSION

BB435111.1 GI:9274838

VERSION

BB435111.1 GI:9274838

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus; Mus.
1 (bases 1 to 273)

REFERENCE

1 (bases 1 to 273)

AUTHORS

Kanno, H., Aizawa, K., Akahira, S., Akiyama, J., Arakawa, T.,
Hirozane, T., Horii, F., Ishii, Y., Ishikawa, J., Ishikawa, T., Itoh, M.,
Izawa, M., Kadoya, K., Kagawa, I., Kai, C., Kawai, J., Kikuchi, N.,
Kiyosawa, H., Kojima, Y., Kondo, S., Koya, S., Kurihara, C.,
Kusakabe, M., Matsuyama, T., Miki, R., Mizuno, Y., Nakamura, M., Oda, H.,
Okazaki, Y., Ono, T., Owa, C., Saito, H., Sakai, C., Sato, K.,
Shibata, K., Shibata, Y., Shigemoto, Y., Shingawa, A., Shiraki, T.,
Sogabe, Y., Sugahara, Y., Suzuki, H., Suzuki, H., Tagawa, A.,
Takahashi, F., Tomimaga, N., Toya, T., Tsunoda, Y., Watanabe, S.,
Watanabe, S., Yamamura, T., Yamana, I., Yano, R., Yasunishi, A.,
Yokota, T., Yoshida, K., Yoshiki, A., Yoshino, M., Muramatsu, M. and
Hayashizaki, Y.

TITLE

RIKEN Mouse ESTs (Kanno, H., et al.)

JOURNAL

Unpublished (2000)

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Shinohe-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp
URL: http://genome.gsc.riken.go.jp/
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoka, S.,
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.

Thermoregulation and its application for thermolabile enzymes by
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (12) 520-524 (1998)

Itoh, M., Katsunari, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,
Okazaki, Y. and Hayashizaki, Y.

Automated filtration-based high-throughput plasmid preparation
system. Genome Res. 9 (5), 463-470 (1999)

Carninci, P. and Hayashizaki, Y.
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
19-44 (1999)

Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.

FEATURES

source

ORIGIN

Query Match
Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

ORIGIN

Query Match
Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

ORIGIN

Query Match
Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

ORIGIN

Query Match
Best Local Similarity 52.1%; Score 25; DB 13; Length 894;
Matches 31; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

2 ATGATAGCTGTTACCCGTTGTTACCAATGAAATGCA 42

Fax: 301 838 0208
Email: szhaoc@tigr.org
Clones are derived from the mouse BAC library RPCT-23. For BAC library availability, please contact Pieter de Jong (pieter@redjond.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm) or from Resea ch Genetics (info@resgen.com). BAC end page: http://www.tigr.org/cdb/bac_ends/mouse/bac_end_intro.html
Plate: 413 row: M column: 16
Seg primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers

1..539

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="RPCT-23-413M16"

/sex="Female"

/lab_host="DH10B"

/note="Organ: Kidney/Brain; Vector: pBAC3.6; Site 1: EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcoRI and EcoRI MethyIase. Size selected DNA was cloned into the pBAC3.6 vector at the EcoRI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies)."

ORIGIN

Query Match
Best Local Similarity 50.4%; Score 24.2; DB 28; Length 539;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 4 GATCAGCTGTTACCGCTGTTGATCCATGAATGCGACATG 48
DB 493 GTTCAGATGTTGCAATGTTGACCCATTCATGTCACATCA 449

RESULT 7
CB200568
LOCUS
DEFINITION
A8ENCCOURT 11271837 NICHED_XGC_Tad2 Xenopus laevis cDNA clone
IMAGE:6875681 5', mRNA sequence.
VERSION
CB200568
KEYWORDS
EST.
SOURCE
Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
1 (bases 1 to 903)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Feb 4, 2003 this sequence version replaced gi:28233219.
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: Drs. Donald Brown and Liqun Cai
CDNA Library Preparation: CLONTECH
DNA Sequencing by: The I.M.A.G.E. Consortium (LIML)
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LIML at:
http://image.llnl.gov
Plate: LHC3111 row: h column: 16
High quality sequence stop: 104.
Location/Qualifiers
1..903
/organism="Xenopus laevis"
/mol_type="mRNA"

FEATURES

Location/Qualifiers

1..903

/organism="Xenopus laevis"

/mol_type="mRNA"

ORIGIN

Query Match
Best Local Similarity 50.4%; Score 24.2; DB 14; Length 903;
Matches 32; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 2 ATGATCAGCTGTTACCGCTGTTGATCCATGAATGCGACATG 46
DB 434 ATGATGAGCTCCCTACTCTATTGCTGCATGAAGTCTACATG 478

RESULT 8
CNS01XV/c
LOCUS
DEFINITION
CNS01XV 1032 bp DNA linear GSS 01-SBP-2000
Tetradon nigroviridis genome survey sequence PUC-ori end of clone
217C04 of library G from Tetradon nigroviridis, genomic survey
sequence.
VERSION
AL172228.1 GI:7810285
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetradon nigroviridis
Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Acanthopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetraodontiformes;
1
Roest Crolius H., Jallion O., Dasilva C., Bouneau L., Fisher C.,
Bernot A., Fizames C., Wincker P., Brothier P., Quetier F.,
Saurin W. and Weissenbach J.
Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
Nuc. Genet. 25 (2), 235-238 (2000)
20296633
10835645

FEATURES

Location/Qualifiers

1..1032

/organism="Tetradon nigroviridis"

/mol_type="genomic DNA"

/db_xref="taxon:99863"

/clone="217C04"

JOURNAL COMMENT

Unpublished (2002)
On Jan 18, 2002 this sequence version replaced gi:18245233.
Contact: Tadao Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsbini@genes.nig.ac.jp.

FEATURES

source

1..600
Location/Qualifiers
/organism="Caenorhabditis elegans"
/mol_type="mRNA"
/strain="N2"
/db_xref="taxon:6239"
/clone="YK1040a11"
/sex="hermaphrodite"
/tissue type="whole animal"
/dev stage="L1"
/clone_lib="unpublished oligo-capped cDNA library, C. elegans L1 stage"

ORIGIN

Query Match 50.0%; Score 24; DB 12; Length 600;
Best Local Similarity 68.8%; Pred. No. 2.2e+02;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY 1 TATGATCAGCTGGTACCCGGTGTGTACCCATGAATGCGACATGCA 48
299 TATATTCACCGATATCCATATATATATATCATCAAAATGCAATTCGA 346

RESULT 12

AL589140/c

LOCUS AL589140 610 bp mRNA linear EST 02-WAR-2001
DEFINITION ROS083H01, mRNA sequence.
ACCESSION AL589140
VERSION AL589140.1 GI:13194174
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 610)
Murray, P.
BP Chicken Brain Library
Unpublished (2001)
Contact: Frazer Murray
Dept. Genomics and Bioinformatics
Roslin Institute
Roslin, Midlothian, EH25 9PS, UK
Tel: +44 (0)131 527 4200
Fax: +44 (0)131 440 0434
Email: frazer.murray@bbsrc.ac.uk
GCGGCCCTTTTCTTTTCTTTT 3' Poly A RNA purchased from Clontech
(*6854-)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENTFEATURES
source

Seq primer: T7.
Location/Qualifiers
1..610
/organism="Gallus gallus"
/mol_type="mRNA"
/db_xref="taxon:9031"
/clone="ROS083H01"
/tissue type="Brain"
/dev stage="unknown"
/lab host="DH10B"
/clone_lib="BP Chicken Brain Library"
/note="Vector: pSPORT1, Site 1: NotI, Site 2: SalI; Cloned
unidirectionally. Primer: Oligo dt. 5' adaptor sequence:
5'-TCGACCTCAG 3', 3' adaptor sequence: 5'
GCGGCCCTTTTCTTTTCTTTT 3' Poly A RNA purchased from

ORIGIN

Clontech (*6854-1)"

Query Match 50.0%; Score 24; DB 9; Length 610;
Best Local Similarity 68.8%; Pred. No. 2.2e+02;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY

1 TATGATCAGCTGGTACCCGGTGTGTACCCATGAATGCGACATGCA 48
397 TACATGATCTGTATCTGTGTGTACATGAATGCGACATGCA 350

RESULT 13

BU294276/c

LOCUS BU294276 766 bp mRNA linear EST 27-NOV-2002
DEFINITION 603065963F1 CSROCHN55 Gallus gallus cDNA clone CHEST590n16 5', mRNA
sequence.
ACCESSION BU294276
VERSION BU294276.1 GI:25743899
KEYWORDS EST
SOURCE Gallus gallus (chicken)
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
1 (bases 1 to 766)
Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E.,
Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
A Comprehensive Collection of Chicken cDNAs
Curr. Biol. 12 (22), 1965-1969 (2002)
22355534
12445392
Contact: Simon Hubbard
Department of Biomolecular Sciences
University of Manchester Institute of Science and Technology
PO Box 88, Manchester, M60 1QD, UK
Tel: 01612008930
Fax: 01612360409
Email: Simon.Hubbard@umist.ac.uk.

REFERENCE

AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

COMMENT

FEATURES
source

Location/Qualifiers
1..766
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="Compton Line 151"
/db_xref="taxon:9031"
/clone="CHEST590n16"
/sex="Female"
/dev stage="adult"
/lab host="DH10B"
/clone_lib="CSROCHN55"
/note="Organ: kidney + adrenal; Vector: pBluescript II
KS(+), Site 1: EcoRI, Site 2: NotI; This normalized
library was constructed from 1 million independent clones.
cDNA synthesis was initiated using an oligo(dT) primer.
Using methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
phasescript sites of a custom modified MCS of the
pBluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldi et al., Genome Research 6
(1996) 791, except that a significantly longer
reannealing hybridization was used."

ORIGIN

Query Match 50.0%; Score 24; DB 13; Length 766;
Best Local Similarity 68.8%; Pred. No. 2.4e+02;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

OY

1 TATGATCAGCTGGTACCCGGTGTGTACCCATGAATGCGACATGCA 48

Db 163 TACAATGATCTGTTATCTGTGTGTTTACAGTAAATAGCCACACA 136

RESULT 14	
LOCUS	B0252275/c
DEFINITION	B0252275/c
ACCESSION	B0252275
VERSION	603748238p1
KEYWORDS	CSEQCHN38 Gallus gallus CDNA clone CHEST556110 5', mRNA sequence.
SOURCE	B0252275
ORGANISM	B0252275.1 GI:25505294
REFERENCE	EST.
AUTHORS	Gallus gallus (chicken)
TITLE	Eukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
JOURNAL	1 (bases 1 to 780)
MEDLINE	Boardman,P.E., Sanz-Ezquerro,J., Overton,I.M., Burt,D.W., Bosch,E., Pong,W.T., Tickle,C., Brown,W.R., Wilson,S.A. and Hubbard,S.J.
PUBMED	A Comprehensive Collection of Chicken CDNA
COMMENT	Curr. Biol. 12 (22), 1965-1969 (2002)
	12335534
	12445392
	Contact: Simon Hubbard
	Department of Biomedical Sciences
	University of Manchester Institute of Science and Technology
	(UMIST)
	PO Box 88, Manchester, M60 1QD, UK
	Tel: 01612008930
	Fax: 01612360409
	Email: Simon.Hubbard@umist.ac.uk.

FEATURES
Source

```

Location/Qualifiers
1..780
/organism="Gallus gallus"
/mol_type="mRNA"
/strain="White Leghorn, Hisex"
/db_xref="taxon:9031"
/clone="CEST656110"
/dev_stage="36"
/lab_host="DH108"
/clone_lib="CSECHN38"
/notes="Organ: Limbs; Vector: pluescript II KS(+), Site:1;
EcoRI; Site_2: NotI. This normalized library was
constructed from 1 million independent clones. CDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was bluntend, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9232 and Bonaldo et al., Genome Research 6
(1996): 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

Query Match 13: Score 24; DB 13; length 780;
Best Local Similarity 68.8%; Pred. No. 2.4e+02;
Matches 33; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

[illegible]

KEYWORDS	EST.
SOURCE	Gallus gallus (chicken)
ORGANISM	Gallus gallus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
AUTHORS	1 (bases 1 to 791)
TITLE	Boardman, P.E., Sanz-Ezquerro, J., Overton, I.M., Burt, D.W., Bosch, E., Fong, W.T., Tickle, C., Brown, W.R.A., Wilson, S.A. and Hubbard, S.J.
JOURNAL	A Comprehensive Collection of Chicken CDNAs
MEDLINE	Curr. Biol. 12 (22), 1965-1969 (2002)
PUBMED	22335534
COMMENT	12445392
CONTACT	Contact: Simon Hubbard
DEPARTMENT	Department of Biomolecular Sciences
UNIVERSITY	University of Manchester Institute of Science and Technology (UMIST)
PO BOX	PO Box 88, Manchester, M60 1QD, UK
TEL	Tel: 01612008930
FAX	Fax: 016123160409
EMAIL	Email: Simon.Hubbard@umist.ac.uk.
LOCATION/QUALIFIERS	Location/Qualifiers
SOURCE	1..791

FEATURES
Source

```

/organism="Gallus gallus"
/mol_type="mRNA"
/dbname="White Leghorn_Hisex"
/strain="White Leghorn_Hisex"
/db_xref="taxon:9031"
/clone="CHS1399510"
/dev_stage="36"
/lab_host="MDH08"
/clone_lib="CSE0CHN3"
/notes="Organ: head; Vector: pluescript II KS(+); Site_1:
EcoRI; Site_2: NotI. This normalized library was
constructed from 1 million independent clones. cDNA
synthesis was initiated using an oligo(dT) primer, using
methylated C in the first strand synthesis reaction.
Following this first strand reaction, double-stranded cDNA
was blunted, ligated to NotI adapters, digested with
EcoRI, size-selected, and cloned into the NotI and EcoRI
compatible sites of a custom modified MCS of the
pluescript (KS+) vector. The library was normalized in 2
rounds using conditions adapted from Soares et al., PNAS
(1994) 91: 9228-9233 and Bonaldo et al., Genome Research
(1996) 791, except that a significantly longer
reannealing hybridization was used."

```

ORIGIN

Query Match	50.0%;	Score 24;	DB 13;	Length 791;
Best Local Similarity	68.8%;	Pred. No. 2.4e+02;		
Matches 33;	Conservative 0;	Mismatches 15;	Indels 0;	Gaps 0;

Qy 1 TATGATCAGTGGTTACCCGTGTTGTTACCATTGAATAAGCACATGCA 48
Db 369 TACATGATCTGGTTATCTGTGTTTACAGTAATAATAGCCCACACA 322

Search completed: March 16, 2004, 04:29:00
Job time : 242.832 Secs

RESULT	15
BUI324056/c	
LOCUS	
DEFINITION	BUI324056 791 bp mRNA EST-28-NOV-2002
ACCESSION	603499343cF1 CSEQCHN63 Gallus gallus CDNA clone ChnST399g10 5', mRNA sequence.
VERSION	BUI324056 BUI324056 . BUI324056.1 GI:25832057

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgcctaacagatctaa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ov:*
8: gb_ov:*
9: gb_ov:*
10: gb_ov:*
11: gb_ov:*
12: gb_ov:*
13: gb_ov:*
14: gb_ov:*
15: gb_ov:*
16: gb_ov:*
17: gb_ov:*
18: gb_ov:*
19: gb_ov:*
20: gb_ov:*
21: gb_ov:*
22: gb_ov:*
23: gb_ov:*
24: gb_ov:*
25: gb_ov:*
26: gb_ov:*
27: gb_ov:*
28: gb_ov:*
29: gb_ov:*
30: gb_ov:*
31: gb_ov:*
32: gb_ov:*
33: gb_ov:*
34: gb_ov:*
35: gb_ov:*
36: gb_ov:*
37: gb_ov:*
38: gb_ov:*
39: gb_ov:*
40: gb_ov:*
41: gb_ov:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	404.8	88.8	456	144909	144909 Sequence 58
2	404.8	88.8	2067	144909	U43280 Salmonella
3	396.8	87.0	5103	144909	AJ002301 Salmonella
4	396.8	87.0	22411	144909	AE008749 Salmonella
5	395.2	86.7	254050	144909	AE027269 Salmonella
6	395.2	86.7	301983	144909	AE016840 Salmonella
7	389.4	85.4	1048	144909	AE00514 Salmonella
8	315.2	69.1	361	144908	AE00514 Salmonella
9	275.2	60.4	2889	144908	AE00514 Salmonella
10	232.4	51.0	2920	144908	AE00514 Salmonella
11	229	50.2	230	144908	AE00514 Salmonella
12	217.6	47.7	4680	144908	AE00514 Salmonella
13	217.6	47.7	10346	144908	AE00514 Salmonella
14	217.6	47.7	15047	144908	AE00514 Salmonella
15	216	47.4	456	144908	AE00514 Salmonella
16	216	47.4	648	144908	AE00514 Salmonella
17	212.6	46.6	306358	144908	AE00514 Salmonella
18	209.4	45.9	1711	144908	AE00514 Salmonella
19	209.4	45.9	10190	144908	AE00514 Salmonella
20	209.4	45.9	327773	144908	AE00514 Salmonella
21	184	40.4	2883	144908	AE00514 Salmonella
22	155.2	34.0	10370	144908	AE00514 Salmonella
23	155.2	34.0	292504	144908	AE00514 Salmonella
24	143.2	31.4	437	144908	AE00514 Salmonella
25	143.2	31.4	1212	144908	AE00514 Salmonella
26	143.2	31.4	19201	144908	AE00514 Salmonella
27	143.2	31.4	72	144908	AE00514 Salmonella
28	143.2	31.4	8.6	144908	AE00514 Salmonella
29	143.2	31.4	2000	144908	AE00514 Salmonella
30	143.2	31.4	6669	144908	AE00514 Salmonella
31	143.2	31.4	1425	144908	AE00514 Salmonella
32	143.2	31.4	302156	144908	AE00514 Salmonella
33	143.2	31.4	7.9	144908	AE00514 Salmonella
34	143.2	31.4	301214	144908	AE00514 Salmonella
35	143.2	31.4	7.8	144908	AE00514 Salmonella
36	143.2	31.4	7.7	144908	AE00514 Salmonella
37	143.2	31.4	230029	144908	AE00514 Salmonella
38	143.2	31.4	547	144908	AE00514 Salmonella
39	143.2	31.4	7.5	144908	AE00514 Salmonella
40	143.2	31.4	151828	144908	AE00514 Salmonella
41	143.2	31.4	7.5	144908	AE00514 Salmonella
42	143.2	31.4	1803	144908	AE00514 Salmonella
43	143.2	31.4	1803	144908	AE00514 Salmonella
44	143.2	31.4	104680	144908	AE00514 Salmonella
45	143.2	31.4	7.5	144908	AE00514 Salmonella

ALIGNMENTS

RESULT 1
LOCUS 144909
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
Doran,J.V., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmomella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

Pred. No. is the number of results predicted by chance to have a

FEATURES

Location/Qualifiers
1..456
/organism="unknown"

ORIGIN
/mol_type="unasigned DNA"

Query Match

Best Local Similarity 93.0%; Score 404.8; DB 6; Length 456;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACGCAATCTAGTTTCTGGAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACGCAATCTAGTTTCTGGAGTCTCTGCT 60
QY 61 GCGGTCTTCCACATGAGGCGGCGGATCATATACGCGCGGCAATAGTTCGCGC 120
DB 61 GCGGTCTTCCACATGAGGCGGCGGATCATATACGCGCGGCAATAGTTCGCGC 120
QY 121 CCGGACTCAAGCTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTGAGCAATTTATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
QY 181 AGCGATCCCGTAAATCTGAAAGCACTTACCGAGCGGTTATGTAAAGCGCGCGAT 240
DB 181 AGCGATCCCGTAAATCTGAAAGCACTTACCGAGCGGTTATGTAAAGCGCGCGAT 240
QY 241 GTAGGCGAGGTCGGATTAATAGTACTGATTAATGAACTGAGTGTTCAGAAATAT 300
DB 241 GTAGGCGAGGTCGGATTAATAGTACTGATTAATGAACTGAGTGTTCAGAAATAT 300
QY 301 GCCACATGACCACTGAGCAATGAAATCCGATTTATCTGCGCCAAATACGCGCT 360
DB 301 GCCACATGACCACTGAGCAATGAAATCCGATTTATCTGCGCCAAATACGCGCT 360
QY 361 AATAACGCGCGCTGTTAATATGATCAGCTGTTACCGGTGTTACCATGAATG 420
DB 361 AATAACGCGCGCTGTTAATATGATCAGCTGTTACCGGTGTTACCATGAATG 420
QY 421 GCACATGCAACACCGCAAGCGCTAACCACTATTAA 456
DB 421 GCTTTGGCAACACCGCAAGCGCTAACCACTATTAA 456

RESULT 2

LOCUS SEU43280

DEFINITION

Salmonella enteritidis agfBAC operon; fimbria-like protein precursor (agfB), thin aggregative fimbriae precursor (agfA), and agfC (agfC) genes, complete cds.

ACCESSION

U43280.1 GI:1184712

VERSION

U43280.1 GI:1184712

KEYWORDS

U43280.1 GI:1184712

SOURCE

U43280.1 GI:1184712

ORGANISM

U43280.1 GI:1184712

REFERENCE

U43280.1 GI:1184712

AUTHORS

U43280.1 GI:1184712

TITLE

U43280.1 GI:1184712

JOURNAL

U43280.1 GI:1184712

MEDLINE

U43280.1 GI:1184712

PUBMED

U43280.1 GI:1184712

REFERENCE

U43280.1 GI:1184712

TITLE

U43280.1 GI:1184712

JOURNAL

U43280.1 GI:1184712

MEDLINE

U43280.1 GI:1184712

PUBMED

U43280.1 GI:1184712

REFERENCE

U43280.1 GI:1184712

AUTHORS

Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg., Victoria, BC V8W 3P6, Canada

FEATURES

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

SOURCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

PUBMED

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

REFERENCE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

TITLE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

JOURNAL

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

MEDLINE

Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"

	Matches	42;	Conservative	0;	Mismatches	32;	Indels	0;	Gaps	0;
QY	1	ATGAAACCTTTTAAAGGTGGCAGCATTCCGACGATCGTAGTTCTTGGCAGTGCCTCGCT	60							
Db	1193	ATGAAACCTTTTAAAGGTGGCAGCATTCCGACGAAATCGTAGTTCTTGGCAGTGCCTCGCT	1253							
QY	61	GCGCGTCTTCACAAATGGGGCGCGCGGTAATCATTAACGGCGGGCGCAATAGTTCGGGC	120							
Db	1253	GCGCGTCTTCACAAATGGGGCGCGCGGTAATCATTAACGGCGGGCGCAATAGTTCGGGC	1312							
QY	121	CCGCACTCAACGTTAGCAATTTATCAGTACGGTTCCGCTAACGCTGGCCTGCTCGCA	180							
Db	1313	CCGCACTCAACGTTAGCAATTTATCAGTACGGTTCCGCTAACGCTGGCCTGCTCGCA	1372							
QY	181	AGCGATGCCCGTAATCTGAAACGACATTTACCCAGACGGGTTATGTTAACGGCGCCGAT	240							
Db	1373	AGCGATGCCCGTAATCTGAAACGACATTTACCCAGACGGGTTATGTTAACGGCGCCGAT	1433							
QY	241	GTAAGCCAGGGTCGGATATATGTAATGTAATGTAAGTACTCAGAAATGTTTCAGAAATAT	300							
Db	1433	GTAAGCCAGGGTCGGATATATGTAATGTAATGTAAGTACTCAGAAATGTTTCAGAAATAT	1492							
QY	301	GCCACCATCGACCAAGTGGACGCTAAATCCGATATTAATCTGTCGGCAATACGGCGGT	360							
Db	1493	GCCACCATCGACCAAGTGGACGCTAAATCCGATATTAATCTGTCGGCAATACGGCGGT	1552							
QY	361	AATAACGCGCGCTGGTATTAATTAATGATCAGCTGTAAACCGTGTGTTAAACCATGAATG	420							
Db	1553	AATAACGCGCGCTGGTATTAATTAATGATCAGCTGTAAACCGTGTGTTAAACCATGAATG	1612							
QY	421	GCAATGCAACAAACGCCACGGCTTAACCAAGTTTAA	456							
Db	1613	GCTTTTGGCAACAAACGCCACGGCTTAACCAAGTTTAA	1648							

```

RESULT 3
STAJ2301
LOCUS          STAJ2301          5103 bp    DNA          linear   BCT 15-NOV-2000
DEFINITION    Salmonella typhimurium cs9G, cs9F, cs9E, cs9D, cs9B, cs9A, and &
              cs9C genes.
ACCESSION     AJ002301
VERSION       AJ002301.1  GI:2739232
KEYWORDS
SOURCE
ORGANISM      Salmonella typhimurium
               Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
               Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS       Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Normark, S.
TITLE         Culi fibers are highly conserved between Salmonella typhimurium
               and Escherichia coli with respect to operon structure and
               regulation
JOURNAL       J. Bacteriol. 180 (3), 722-731 (1998)
MEDLINE       98117058
PUBMED        9457880
REFERENCE     2 (bases 1 to 5103)
AUTHORS       Romling, U.
TITLE         Direct Submision
JOURNAL       Submitted (22-OCT-1997) Romling U., Department of Bacteriology,
               Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN
FEATURES
SOURCE        location/Qualifiers
               1..5103
               /organism="Salmonella typhimurium"
               /mol_type="genomic DNA"
               /strain="SR-11"
               /db_xref="taxon:602"
               <1..322
               /note="ORF199a"
               /codon_start=2
               /transl_table=11
               /product="hypothetical protein"
               /protein_id="CA05311.1"

```

[illegible]

```

/translation="MSVKKKNIPIALGICAFHSANGQQTVCQGVTHFNGALVEPL
CDISTHMENT"
/db_xref="SMISS-PROT:054295"
/protein_id="CNA05319.1"
/db_xref="GI:2739241"
/translation="CNA05319.1"
/origin

```

Query Match	Best Local Similarity	91.9%	Score 396.8	DB 1	Length 5103
Matches 419	Conservative	0	Mismatches 37	Indels 0	Gaps
Qy	1	ATGAAACCTTTTAAAGTGGCAGCATTTGGCAGCAATGCTAGTTCTTCTGGCAGTGTCTGGCT	6		
Db	3966	ATGAAACCTTTTAAAGTGGCAGCATTTGGCAGCAATGCTAGTTCTTCTGGCAGTGTCTGGCT	6		
Qy	61	GGCGTCGTTCCACAAATGGGGCGGGCGGTAAATCAATACGGGGGGGCAATAGTTCGGC	12		
Db	4026	GGCGTCGTTCCACAAATGGGGCGGGCGGTAAATCAATACGGGGGGGCAATAGTTCGGC	12		
Qy	121	CCGAGCTCAACGTTGAGCATTTATCAGTACGGTTCCGTTACCTGCGCTTCTGGCA	18		
Db	4086	CCGAGTTTCCACGTTGAGCATTTATCAGTACGGTTCCGTTACCTGCGCTTCTGGCA	18		
Qy	181	AGCGATGCCCGTAAATCTGAAACGACATTCATCCAGACGGTATAGTAAACGGCGCGAT	24		
Db	4146	AGCGATGCCCGTAAATCTGAAACGACATTCATCCAGACGGTATAGTAAACGGCGCGAT	24		
Qy	241	GTAGCCGAGGGTGGGATTAATGATCTATTGAACTGATCGATCAGATAGTCTTCAAAATAT	30		
Db	4206	GTAGCCGAGGGTGGGATTAATGATCTATTGAACTGATCGATCAGATAGTCTTCAAAATAT	30		
Qy	301	GCACCATCATGACACAGTGAAGCGCTAAATCTCCGATTTACTGTGGCAATACGGCGGT	36		
Db	4266	GCACCATCATGACACAGTGAAGCGCTAAATCTCCGATTTACTGTGGCAATACGGCGGT	36		
Qy	361	AATPACGCGCGCTGTGTTAATTATGATCAGCTGGTACCCTGTTGTTACCATGAAATG	420		
Db	4326	AATPACGCGCGCTGTGTTAATTATGATCAGCTGGTACCCTGTTGTTACCATGAAATG	420		
Qy	421	GCACATGCAACCAACGCCACGCGCTAACCGATTTAA	456		
Db	4386	GGTTTTCGCAACACGCCACGCGCTAACCGATTTAA	4421		
RESULT 4	AE008749	22411 bp	DNA	linear	BCT 23-APR-2003
LOCUS	Salmonella typhimurium LT2, section 53 of 220 of the complete genome.				
ACCESSION	AE008749	AE006468			
VERSION	AE008749.1	GI:16419641			
KEYWORDS					
SOURCE					
ORGANISM	Salmonella typhimurium LT2				
REFERENCE	Salmonella typhimurium LT2				
AUTHORS	Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.				
	1 (bases 1 to 22411)				
	McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,				
	Latreille, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,				
	Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,				
	Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,				
	Stromberg, T., Nhan, M., Waterston, R. and Wilson, R. K.				
TITLE	Complete genome sequence of Salmonella enterica serovar Typhimurium LT2				
JOURNAL	Nature 413 (6856), 852-856 (2001)				
MEDLINE	21534948				
REFERENCE	11677609				
AUTHORS	2 (bases 1 to 22411)				
CONSTRM					
TITLE	The Salmonella typhimurium Genome Sequencing Project				
JOURNAL	Direct Submission				
	Submitted (29-MAR-2001) Genbank				

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 A143283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER: <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
ReguionDB;

http://kinich.cifn.unam.mx:8850/db/reguiondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m3 subclone.

FEATURES

SOURCE

Location/Qualifiers
1..22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
434..1308
/gene="phoH"
/note="synonym: STM1126"
434..439
/gene="phoH"
/note="putative RBS for phoH; ReguionDB:STMS1H001398"
454..1308
/gene="phoH"
/note="similar to E. coli Phob-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=1
/product="Phob-dependent ATP-binding pho regulon
component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="MCANRVQENAGDVYIQLKLSNGRQKAVIKARREARKVLRRDS
RSHKQREESVTSLVOMGVGAIGMARDSTSVKANEQAHYLNADSKOIFAT
GEACCGKWTISAAKAAEALIKHVERIIVTRPVLOADELGFLLPGDIAEKAPYFRPV
YDVLLKRLGASFMQYCLRPETIKVEIAPFAMRGRTENAVIILDEAQNVTAAQMKP
LTRGENTVTVIVNDITQCCLPRGVRGSLDARFREDEMGIVYHNRKDDCVASALC
QRTLHAYS"
complement(1414..2303)
/gene="STM1127"
complement(1414..2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEPPRLKPKGLDITLGMOKSLTPRASQRIAYIILAFPROVTO
SSIALDRTOAGATVIRCRITGKGFDFDKNDLAIELATTSDDSSPLDAVEB
SDMAHIGLKQNTISNVLTLLDMQVLLGVDAIRHCHSYVIFGVSGGTTAD
MKHLMRIGLRGDAVSNHFMVQATLLKAGDVMGVSHSGTSPETVHSLRLAQGA

gene

RBS

CDS

RBS

gene

CDS

RBS

gene

CDS

TTVATHNIGSPICEADFCILINGNRQMLQGSIGTKAQLFVFDLLYTLNQSPE
QARSKIRTNALDMTK"
complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; ReguionDB:STMS1H001398"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC6702.1); Blastp hit to AAC6702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSGIVNYLVLFGLAMLVGVYRSRROKTDADYPRGGR
VPGWAGVSPATLSSITMSIPAKAFTSPMTIIGQYLAIALIPVFYIPEPRK
LKVTSATVELEARDVRCRLFRASMPHLFHIGRAIITFTLVTLARPIAIDPVLV
LISVCIITVMGGBEVWTDVIOGLLSALILFVLCIKVOGGIDEFTVQQA
DKFPPAOPFMSWTSTVPMVIMIGFLPANTIOQFPASDVORVYVTSIEBTKTLT
NAKLAVIIPVFPAIGSALFVYQOQLPAGCNTGILPLFVTEPMPVIGAILIA
AIFPAOSISISSISCSFSDIYORLSHKRTPEMRKIAKLVIIVAGLSAAS
VMIYMADESEIMDAFNSLIGIMGSPMTGLMGLFFPRANSAVLGIISIVTVLGA
RYATDLPFFYGVGSLSVYISGVIAPLAPAPPLTIDEBPKVTL"
complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; ReguionDB:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSILARLEOSVHENGGLIVSCOPVPSGPMDCPEIVAAQQAAS
AGAAVAVRIGENIRTVRPHLSVPIIGILRDLGSRVTRIPYODDALQAGDII
AFDASFRSREVDISLIRIRHGLMLMADCSYTBESICGQKITEFRTLLSGTTC
ITVPEPDLAVTQLSHAGCRVLAEGRTPALAAALBHGAAVTVGSATRIEHTCQ
WFSHAVXR"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; ReguionDB:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGMQKNPFKQMTIIMLCSSVAITTSQYATLLPDIEPLKNGTG
AIDNNGVIVGLGAGTSWYKIDLKQKHMERIKSPFGARGESVSLFDELIVTG
GVGNKSESPLOYSDVYKSPVNTQKQDITSPVLTGHTGKLEBTVNLITGVN
EHIIFDKYFIDIAAADADESEKNKYIVFVKRQVLEFNKIVFLYNAKENTYKAGSL
PGAGTAGSSVMGNFLMLINGELKPLRFVDVYRAMWDDKLTMLKNSQLPSPGBO
QOGSLGAFSGYSHGVTLVGGANPFAKQNTYNGKYSHEGINKKRDVYVGIINH
WQYMGKKKQPLGCVSYSGDEVFLIGENAKKRPVSYSFTMRDNLILK"
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; ReguionDB:STMS1H001402"

CDS

6805..7497
/gene="STM1131"
/note="similar to E. coli orf. hypothetical protein
identity in aa 4 - 239"
/codon_start=1
/transl_table=11
/product="putative outer membrane protein"
/protein_id="AA020062.1"
/db_xref="GI:16419647"
/translation="MKINKYLGVSVFIAFSSSYLAQATIDYRHEVADTRINKORAI
IEKLPNIGIYVDASVSGVDGEODKHSIDIVANAIEGVSINKVNDVLPQGI
FESGPDISIKYPIRGYNFDSGVWAGYRDPYRKRANSDSEKTRFPTIYGFY
DELKLEYNFTWMSDQIKPDNKTNYENHVALAMKLNKSTPYVEGVAVRNTDER
QTRYRVGLQYH"
/gene="STM1132"
7768..7773
/gene="STM1132"
/note="putative RBS for STM1132; RegulonDB:STMS1H001403"
7780..9060
/gene="STM1132"
/note="similar to E. coli putative transport protein
identity in aa 25 - 418"
/codon_start=1
/transl_table=11
/product="putative sugar transport protein"

Query Match 87.0%; Score 396.8; DB 1; Length 22411;
Best Local Similarity 91.9%; Pred. No. 1.3e-101;
Matches 419; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY	1	ATGAACTTTTAAAGTGGCAGATTGGCAGCAATGTAATCTTGGCAGTCTCGGCT	60
DB	17769	ATGAACTTTTAAAGTGGCAGATTGGCAGCAATGTAATCTTGGCAGTCTCGGCT	17828
QY	61	GGCGTCTTCCACAAATGAGGCGGCGGCGGTAATCAATAGCGGCGCAATGTTCCGCG	120
DB	17829	GGCGTCTTCCACAAATGAGGCGGCGGCGGTAATCAATAGCGGCGCAATGTTCCGCG	17888
QY	121	CCGGACTCAAGTGTGAGCATTTATGATGATCGGTTACGCTGCGCTGCTGCA	180
DB	17889	CCGGACTCAAGTGTGAGCATTTATGATGATCGGTTACGCTGCGCTGCTGCA	17948
QY	181	AGGATGCGCGTAAATCTGAAAACCAATACCAAGCGGTTATGTAACGCGCGCAT	240
DB	17949	AGGATGCGCGTAAATCTGAAAACCAATACCAAGCGGTTATGTAACGCGCGCAT	18008
QY	241	GTAAGCGCAGGTCGGGTAATGATCTTGAATCACTGACTCAAGATGTTTCAAGAAATAT	300
DB	18009	GTAAGCGCAGGTCGGGTAATGATCTTGAATCACTGACTCAAGATGTTTCAAGAAATAT	18068
QY	301	GGCACCATCGACCAATGGAACGCTTAAACCTCCGATATCTGTCGCGCAATACGCGCGT	360
DB	18069	GGCACCATCGACCAATGGAACGCTTAAACCTCCGATATCTGTCGCGCAATACGCGCGT	18128
QY	361	AATAACGCCGCGCTGTTAATATGATCACTGCTGTTACCGCTGTTTCAAGCAATATG	420
DB	18129	AATAACGCCGCGCTGTTAATATGATCACTGCTGTTACCGCTGTTTCAAGCAATATG	18188
QY	421	GCAATCGCAAAACAGCGCCACGCTTAAACCAATATTA	456
DB	18189	GCTTTGGCAACAGCGCCACGCTTAAACCAATATTA	18224

RESULT 5
AL627269 254050 bp DNA linear BCT 04-JUN-2003
LOCUS Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
DEFINITION complete chromosome, segment 5/20.
ACCESSION AL627269 AL513382
VERSION AL627269.1 GI:16502231

KEYWORDS

Salmonella enterica subsp. enterica serovar Typhi

SOURCE

Salmonella enterica subsp. enterica serovar Typhi

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

REFERENCE

1 (bases 1 to 254050)
Parhill, J., Dougan, G., James, K. D., Thomson, N. R., Pickard, D.,

AUTHORS

Sehbiha, M., Baker, S., Baaham, D., Brooks, K., Chillingworth, T.,

TITLE

Complete genome sequence of a multiple drug resistant Salmonella

JOURNAL

Nature 413 (6858), 848-852 (2001)

MEDLINE

21534947

PUBMED

11677608

REFERENCE

2 (bases 1 to 254050)
Parhill, J.

AUTHORS

Submitted (25-OCT-2001) Submitted on behalf of the Salmonella

TITLE

Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

JOURNAL

Hinxton, Cambridge CB10 1SA, UK

COMMENT

Notes: Details of S. Typhi sequencing at the Sanger Centre are available

FEATURES

on the world wide Web.
(URL, http://www.sanger.ac.uk/Projects/S_typhi/).

source

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

FEATURES

1..254050
Location/Qualifiers

gene /note="PS00142 Neutral zinc metalloproteinases, zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82759"
/db_xref="Swiss-Prot:Q82759"
/translation="MYFFPKALFQDLDEPAHEFTFOQLRITGTPLLEALVRQKVP
KPTCMGLTFKPLGLAAGLDDECDIPALGAMGFSLEIGTVPRPOGDKPLFR
LVDAEGLNRKGFNNLGVNLYENYKKAFFGIIIGINRKNDDPVENKXDYLIME
KVYAVGYIAINISPNRGLRTLOYGDLDDLITAJIKKNDLQVHHKRVPAVKI
APDCBEELIQAADSLRNITDGIATTTTIDRSIVQMKKCCQGTGLSGRPLOKST
EIIIRLSLELKGQPLIGVGIDSVIAREKIAAGATLVQISGFIKGPPLIKEIYV
HI"
misc_feature 3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHodehase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q82758"
/translation="MMRIKPDNNRWYDEEDHMMMLDLANGMLFRSFRSKMLTPPA
FCTGFCVDAAALYFSPEEKCRDELTEKORALVLNALVALRYLKQPMRSMHFAV
GEMTPTGDAASVMSDPAQVNLVYEPBENALCLIAQPGVIAGRTVQLDALK
IMNRLKPOVHCHSFSLBOAV"
CDS complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/db_xref="SPTREMBL:Q82757"
/translation="MATLSRLFIHPVSMRGIGLTHALADISGLAFRIEMVTSDEGT
FITARQPMWRFTPSPLDGLHTLPAKSGSALVRFDTQDAPTEWGNHFAFARA
PTAINOMLSGFSRDLVQCMGPOLTRKVNHNVAVPGFAGVYLLNLSLSDLOO
RCPAGVMEQRPMLVYSGVAWEKDSKULRIGDVFEDVYKPCSRICFTTYSPEKQ
KHPSCEPLATIQARTNODNGDVPCQULIARNSGVIRVGEVITLATAKAYGATT
LDGVSVEPKKSDGVTIDMGQFTCGNNQVLLLEQLEQSGIRIYPSGAGICGCRIR
LLEGEVPLKKSAMGDDGTIISCVPTALRLN"
misc_feature complement(4737..4955)
/gene="STY1081"

gene /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein H10116 H10116 SW:YCBY_HAEIN (P44524; P43945) faeta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82756"
/db_xref="SPTREMBL:Q82756"
/translation="MNSLPASTAGLEELIKIELEKAGVCCOVVQGVHRCQTRLI
YQSLMWSRLASRIILPWECKVYSDDLIFGVQAINTEIFNPATRAVHPSGLDITI
RNSQYGMKAKDAIVDAFTRKNLPRPVNDRESPOLRLNWLNKETASIALDSDGLH
LRGDRDRTGLAPIKETLAAAIYMSGQPGTPLDPMCGSGTLLIEAMATDRAPIGL
HRGHMGSFGMAOHDEAIWQEKVKAOTRARKGLAEYSHSHPGSDSARVIERASNNR
RAGIGELITFEVQVQAQNSLPKPGYQTVISNPVGBRLSDPALIATSLGRTYK
NORGWMLSPFSASPDLISQLRADQFPAKNSPLDQKNTYHIAETADSKATYA
EDIANRRLKULKEKMARQEGIECTVLDLDETVAAVDREDAVIOBAPKTV
DAQKARQLFDIIAATLSVGIIPNNKVLITREKQKNQYQKMSKEGFELEBYNA
RLWNLITDYDITGLFLDRIIARMLGMSKQPLNFSTYGSASVHAGLGAGANTT
VDMSTRYLEMAERNLRNLNGSRAHRLIQDDCGLMLEANEOPLIFIDPPTFENSKR
MESEFDVORQHVALLKDLKRLKRGKGTIMSNRGRPMDEGLAEGLTAEITQKT
LSDFPARNROIHNKMLIRAA"
misc_feature 5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa overlap

Query Match 86.7%; Score 395.2; DB 1; Length 254050;
Best Local Similarity 91.7%; Pred. No. 3.9e-101;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY	1	ATGAACCTTTAAAGNGGACGACATTCGACGACATCGATGTTTGGCAGTCTGGCT	60
Db	88906	ATGAACCTTTAAAGNGGACGACATTCGACGACATCGATGTTTGGCAGTCTGGCT	88965
QY	61	GGCGTCGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTTCGGC	120
Db	88966	GGCGTCGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTTCGGC	89025
QY	121	CCGACCTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGGCTTCTGCA	180
Db	89026	CCGACCTCAACGTTGAGCATTTATCAGTACGGTTCCGTTAACGCTGGCTTCTGCA	89085
QY	181	AGCGATCGCGGTAAATCGAAGACGACATTACCGAGAGCGGTTATGGTAACGGCGCAT	240
Db	89086	AGCGATCGCGGTAAATCGAAGACGACATTACCGAGAGCGGTTATGGTAACGGCGCAT	89145
QY	241	GTAAGCCAGGCGGTGGAATATAGTACTATGAACTGACTGAAATGGTTTCAGAAATAT	300
Db	89146	GTAAGCCAGGCGGTGGAATATAGTACTATGAACTGACTGAAATGGTTTCAGAAATAT	89205

gene
RAUNAATGTPMDLMSILQSMOAGCLAKLEIT"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
/notes="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/transl_table=1
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKYLAHAPYVMAALMTLTGCAWIPAKPLVQAGTAQPIPGV
PVANGSIPOSQPTNGYQPLFEDRPRLNIDTLIVLQENVASKSSAASRDGKT
SFGEDVRYLQGLFSGSRADMSGNSFNGKGAASNTFSGTLTVTDQVLANN
LHVVGKQIAINQCTEEFRFGSVNPRITSGNSVSPSTQVADARIEYVNGYINEQN
MGMLQRFELNLSPM"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
/notes="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="flagellar basal-body rod protein flgC"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLWIAKTGLDAQCTNMVDVIANNLANSSTNGPKQRAVEIL
LYOTIRPGAOSSQTLPSGLQIGTVRPVATERLHSGALSGTNSKQVALKQGF
FOVWLPGTSAYTRDGSFOVDNQGLTAGFOVQPAITIPANLSTIGSDGVSAT
QOQCAAPVQGLMLTFMNDTGLSISGENLYETQSGAPNESTPLNGAGLLYQCY
VETSNVVAEELVMIOVARYEINSAVSTTDMLQKLTQL"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
/notes="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="WDAIITTMGASQTLNQAVTNANLANSSTPGFRAQNLNLRAY
PVGLSLATRTLVASTPGADMTQGLDYTRPLDVALQDDGMLVVOAADGASGTEN
GNIOVGTQGLTIQGHVIGEGPITVPESELTITADGTISALNPEDPNTVAQVPR
LKLVAEGNEVQSDDLFRLLTAQAQERAVLAAADPSIRIMSGVLEGSNVKPEAMT
DMIANARFEMOKVITSVDNEGRANQLLSMS"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
/notes="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSRSQAVSGANAAATLVDVIGNNANSATYGPFSQASPADMA
GSRVGLSVKAVGITQDFTDGTITNGDLVAISQNGFRLVDNSNGSVFYSRNGQFCL

Query Match 86.7%; Score 395.2; DB 1; Length 301983;
Best Local Similarity 91.7%; Pred. No. 3.9e-101;

Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGCGAGTCTGGCT 60
Db 37210 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTCTGCGAGTCTGGCT 37251
QY 61 GGCCTCGTCCACAAATGGGGGGGGGGGTATATCAACGGCGCGCAATAGTTCGGC 120
Db 37250 GGCCTCGTCCACAAATGGGGGGGGGGGTATATCAACGGCGCGCAATAGTTCGGC 37191
QY 121 CCGGACTCAACGTTGAGCATTATTCAGTACGGTCCGCTAACGTCGCTCTGCA 180
Db 37190 CCGGACTCAACGTTGAGCATTATTCAGTACGGTCCGCTAACGTCGCTCTGCA 37131
QY 181 AGCATGCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
Db 37130 AGCATGCCCGTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 37071
QY 241 GTAGCCAGGCTGGGATTAATAGTACTGATGACTGATGCTGCTCAAGAAATAT 300
Db 37070 GTAGCCAGGCTGGGATTAATAGTACTGATGACTGATGCTGCTCAAGAAATAT 37011
QY 301 GCCACATCGACCGAGTGAAGCTAAACTCCGATATTAATCTGCGGCAATACGGCGGT 360
Db 37010 GCCACATCGACCGAGTGAAGCTAAACTCCGATATTAATCTGCGGCAATACGGCGGT 36951
QY 361 AATAGCCGCGGTGTTAATATGATCAGCTGGTTACCCGTGTTTACCAGTAATG 420
Db 36950 AATAGCCGCGGTGTTAATATGATCAGCTGGTTACCCGTGTTTACCAGTAATG 36891
QY 421 GCACATCAAAACGCGCAGCGCTAACCATTAAT 456
Db 36890 GCTTTGGCAACACGCGCAGCGCTAACCATTAAT 36855
RESULT 7
STAGBFA
LOCUS STAGBFA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeifer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
Infect. Immun. 65 (12), 5320-5325 (1997)
98053981
9393832
2 (bases 1 to 1048)
Sukupolvi,S.S.
Direct Submission
Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllynkatu, 20520, FINLAND
location/Qualifiers
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=1
/product="AgfB protein"
/protein_id="CAA04150.1"

```

/db_xref="GI:2275120"
/db_xref="SPTREMBL:O33801"
/translation="MKKLLPMLLTJIEPIATNTYDLARSEYFAVELSKSFNO
AAIGVGTDSARVREHSGKSLISVIGERGNARAVDQGNFAYIEQGNANDA
SISOSAYGNSAISAIIOKSGSKANITGYGKTAVVVQKSHMAIDANTGYGTO
KTAVVVQKSHMAIRVTR"
593..1048
/gene="agfa"
/gene_start=1
/gene_end=1048
/codon_start=1
/codon_end=1048
/product="agfa protein"
/protein_id="CAA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKLLKVAFAALVSGSAVAVGVPMGGGNNHNGGNSGSPST
LSIYVGSANALALOSDKRSETTITGSGNGADVGADNSTIELTONGFENNAT
IDOMNARSDITVGYGNNALVNTASDSVWVRQFNNAPANQYN"
ORIGIN

```

```

Query Match
Best Local Similarity 85.4%; Score 389.4; DB 1; Length 1048;
Matches 414; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCAGATTCAGCAATCGTAGTTCTGCGAGTCTGGCT 60
DB 593 ATGAACCTTTAAAGTGGCAGATTCAGCAATCGTAGTTCTGCGAGTCTGGCT 60
QY 61 GGGCTGCTTCAAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 653 GGGCTGCTTCAAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAAGCTTGAAGTATTAATGAGTGGTGGTGGTGGTGGTGGTGGT 180
DB 713 CCGGACTCAAGTGGTATTAATGAGTGGTGGTGGTGGTGGTGGTGGTGGT 180
QY 181 AGGATGCGCCGCTAAATCTGAACGACATTAACGAGGGGTTATGTAACGGCGCGAT 240
DB 773 AGGATGCGCCGCTAAATCTGAACGACATTAACGAGGGGTTATGTAACGGCGCGAT 240
QY 241 GTAGCGCAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
DB 833 GTAGCGCAGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 300
QY 301 GCCACCATGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
DB 893 GCCACCATGACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360
QY 361 AATAACGCGCGCTGTTAATATGATGAGTGGTGGTGGTGGTGGTGGTGGTGGT 420
DB 953 AATAACGCGCGCTGTTAATATGATGAGTGGTGGTGGTGGTGGTGGTGGTGGT 420
QY 421 GCACATGCAACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
DB 1013 GCACATGCAACGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 480
QY 1013 GGTGGTGGCAGACGACGCGCTAACCAAGTATTA 1047

```

```

RESULT 8
LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION 144908
VERSION 144908.1 GI:2469621
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 361)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

```

```

FEATURES
Source Location/Qualifiers
1..361
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN
Query Match
Best Local Similarity 69.1%; Score 315.2; DB 6; Length 361;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 64 GTGCTTCAACATGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123
DB 1 GTGCTTCAACATGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123
QY 124 GACTCAACGTTGAGCATTATTCAGTACGGTTCGGCTAACGCTGGCTGCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTATTCAGTACGGTTCGGCTAACGCTGGCTGCTGCAAGC 183
QY 184 GATGCCCGTAAATCTGAACGACATTAACCGAGCGGTTATGTTAAACGGCGCATGTA 243
DB 121 GATGCCCGTAAATCTGAACGACATTAACCGAGCGGTTATGTTAAACGGCGCATGTA 243
QY 244 GGGCAGGGTGGGAGTATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 303
DB 181 GGGCAGGGTGGGAGTATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 303
QY 304 ACCATGACACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 363
DB 241 ACCATGACACAGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 363
QY 364 AACGCGCGCTGTTAATTA 383
DB 301 AACGCGCGCTGTTAATCA 320

```

```

RESULT 9
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. fec2 csbA gene, csbA gene and csbD gene.
ACCESSION AU515700
VERSION AU515700.1 GI:31790491
KEYWORDS csbA gene; csbD gene; curlin-csbA protein; nucleation
component of curlin monomers; regulatory protein.
SOURCE
ORGANISM
REFERENCE 1
AUTHORS Zogaj,X., Bokranz,W., Nintz,M., and Rowling,U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family
JOURNAL Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
REFERENCE 2 (bases 1 to 2889)
AUTHORS Rowling,U.
TITLE Direct Submission
JOURNAL Submitted (11-NOV-2002) Rowling U., Microbiology and Tumourbiology
Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
Source Location/Qualifiers
1..2889
/organism="Citrobacter sp. fec2"
/mol_type="genomic DNA"
/isolate="fec2"
/isolation_source="faeces"
/species="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csbD"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

Qy	1	ATGAAACTTTTAAAAAGTGGCAGCAATCCGACCAATCGAGTTTCTGGCAGAGCTTGCT	60
Db	2119	ATGAAACTTTTACAAGTGGCAGCAATTTGACCAATCGAGTTTCTGGCAGAGCTTGCT	2178
Qy	61	GCGCTGTTCCACAATGGGGCGCGCGCGGTAAATCAATACGGCGCGGCAATAGTTCCGGC	120
Db	2179	GGTTCTGTTCCGCAATGGGGCGGCGCGG---TGGCGGCGCGCGGGAGGAGCTCCGGC	2239
Qy	121	CCGGACTCAAGTTGAGCATTTTATCAGTACGGTTCCGCTAACCGTGGCCTTGCTCTGCA	180
Db	2236	CCGGAATCGACCTCGAGCATTTTATCAGTACAGAGATCAATTAACGCGCGCTTGCGCTGCA	2239
Qy	181	AGCATGCGCCGTTAAATCTGAACAGCACCATTACCAGACGGGTTATGGTAAACGGCGCGCAT	240
Db	2296	AGCAGCTGCTTAATCTGATACGACCATTTATCAGAATGGCTTTGGTAAACGGCGCGCAGC	2355
Qy	241	GTAGCGCCAGGGTGGCGATATAGTACTATTGAACTGACTCGAAATGGTTTCAGAAATAT	300
Db	2356	GTGGCGCCAGGGCTCAGATTAACAGCACCATTCGATCGATCAAAACGGGCTTCAAAAACAC	2415
Qy	301	GCCACCAATCGACCACTGGAAACGGCTTAAAACTCCGATATTACTGTCCGCAATACGGCGGT	360
Db	2416	GCCACCAATCGACCACTGGAAACGGGAAAAATTCGACATTTACTGTGAAGCCAGTAATGGTGA	2475
Qy	361	AATAACGCGCGCTGGTTAATTATGATCAGCTGGTATCCCGTGTGTTATCCACATGAATG	420
Db	2476	CATACCGCGGACCTGGTAAACCAACAGCTGCGCTCGCATTTCAAGCTTCTGTGCACTCAGGTT	2535
Qy	421	GCACATGCAACAAACGCCACGGCTAACCACTGATTAA	456
Db	2536	GGTTTGGCAACAACGCCACCGCTAACCACTGATTAA	2571

```
/translation="MKLLKVAAPFAIVSGSALAGVPPWGGNHHGGGSNYGPPSSLS  
IYQYGSNNANALQSDARKSDVITTOHGRNGAVVGQADGSTISLKQTFQNSATID  
QNNAKNADISVTFGGRNGALVNQTSADSNVLIQQVFGNNATANQH"
```



```
/gene="csbg"
complement(133. .966)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62277.1"
/db_xref="GI:1147563"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORFLVAVMLISGCLTAPPKEARPTLMPRAOSYKDLTHLPA
PGKIFVSVYIODETFQKPYPAVSFSTAVQSATMLVTALDSRKFPLESGQLO
NLNBRKTIIPAQNGCTVAIINRIPIOSITANIMGSIIGSYESNKGSGVARYG
IGADTQYQLODIAVNLRVNVSTGEILSSVTSKTIISYEQAGVFRIQRLBGE
VGYTSNEPVMCLMSALETGVIFLINDIDGLMDLONKARONDILVKYHMSVPP
S"
complement(993. .1409)
/gene="csbg"
complement(993. .1409)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translation="MRYKHAVVLLMLISPLSWAGTMTQFPNPNFGNPNNGAFILNS
AOQNSYKDPFYNDFGIEFSPALDFTQALQSOILGGLSLNINTGKRGKRWVTDYIV
DIARDGLOLQNLDRKTGOTSTIQVSGLOKNNSTDF"
complement(1434. .1823)
/gene="csbg"
complement(1434. .1823)
/gene="csbg"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CAA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translation="MKRIKRWIVAAEFLLFAAGNLHAEVEVGLTHTVSSIGHDFY
RASFDMKSDVYTNLTINERPSARMSGITTVQDVIFQFELPLKRDPEKTVVPL
IQTEBALNRQINALLSTGLADPE"
complement(1828. .2478)
/gene="csbg"
complement(1828. .2478)
/gene="csbg"
/notes="necessary for transcription of the csbA operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CAA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="WFNEVSHIGHTLITLTKSLQATALIQHLKQSLATKGLHNTQ
RSJLDISSGSIILDMNEADKLIHYQDTLSRNNNLIKLLNTPBEDPYRDIENNP
HINGVFYSMEDQERVNGLQVLRGECFTQKLASYLITSGNRYNRSBESALLTHEE
KEINLKRIGASNNEIARSLPISENTVYKHLNLPFKIAVNGRTQAVSWANDNLR"
complement(2626)
/notes="transcriptional start site for csbD operon"
3140
/notes="transcriptional start site for csbA operon"
3233. .3688
/gene="csbg"
3233. .3688
/gene="csbg"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
/protein_id="CAA62281.1"
```

```
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKKKLLFMMLTLLIGAPGIAAAGYDLANSYFNAVELSKSFN
QAAIIGQGTNNSAQLRQGSKLAVVAEGSSNRRAIDQTDYNAIYIIDAGSANDA
SISQAGYGNAMTIQKSGNKNANITQVQKTALIVVQSQMAIRVQR"
3729. .4184
/gene="csbg"
3729. .4184
/gene="csbg"
/notes="major subunit"
/codon_start=1
/transl_table=1
/product="curlin/CsgA protein"
/protein_id="CAA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="MKLLKVAIAIAIVFSGSALAGVVPYQGGGNGHGGGNGGPNSE
LNIYQVGGNSALALQTDARNSDLTIHQHGNGADYQGSDDSSIDLTRGFNSAT
LDQNGKNSERTYKQFGGNGAAYDQASNSVNVTVQFGFNATAHQY"
3729. .3788
/gene="csbg"
4243. .4575
/gene="orf"
4243. .4575
/gene="orf"
/codon_start=1
/transl_table=1
/product="putative curli production protein"
/protein_id="CAA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLLLAALSQITNTNTTQGGDVYTIIEPTVLNQSLCVRQL
SLRSGSGQSQTKQEKTLSPANQPIALTKLSINISPDVRKIVIVTVSDGSLHLSQQ
WPSSEKS"

ORIGIN
Query Match 47.7%; Score 217.6; DB 1; Length 4680;
Best Local Similarity 67.3%; Pred. No. 1.2e-50;
Matches 307; Conservative 0; Mismatches 149; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGCGAGCATTCGACATCGAGTTTCTGGCAGTCTCTGGCT 60
Db 3729 ATGAACCTTTAAAGTGGCGAGCATTCGACATTCGATTCCTCGTAGCGCTCGCA 3788

QY 61 GGCCTCGTCCACATGAGGGGCGGCGGCTAATCATACGCGCGCAATAGTTCGGC 120
Db 3789 GGTGTGTTCTCTAGTACGCGGGCGGCTAACCAGGTGTGCGGTAATATAGCGGC 3848

QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGTCGCTTGCTGCA 180
Db 3849 CCAAAATCTAGCTGAACATTACACAGTACGGTGGCGGTAACTTCGACTTCTCGAA 3908

QY 181 AGGATCCCGTAAATCGAAGACATTACCAGAGCGGTTATGGTAACGCGCGCAT 240
Db 3909 ACTATGCCCCGTACTCTGACTTACATTAACCAAGCATGCGCGCGGTAATGTCAGAT 3968

QY 241 GTAGCCAGGGTGGGATTAATAGTACTATTGAACCTGAACGAAATGCTTCAGAAATAT 300
Db 3969 GTTGTCAGGGCTCAGATGACAGCTCAATCGATCGAACCAATGCTTCGTAACGAC 4028

QY 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTACTGTCGCGCAATACGGCGGT 360
Db 4029 GCTACTCTTATCATGATGAACGCAAAATTCGAAATGACGCTTAAACAGTTGCTGCT 4088

QY 361 AATAACCGCGCGCTGTTAATTAATGATCAGCTGGTTACCGGTGTTTACCAGTAATG 420
Db 4089 GCGAACGCTGCTGACAGTTGACAGACTGCATCTAACCTTCGCCAAGTACTCAGGTT 4148

QY 421 GCACATGCAACCAACGCGACGCGTAAACAGTATTA 456
Db 4149 GCGTTTGTAAACGCGACCGCTCATCAGTACTAA 4184
```

```

RESULT 13
AE000205
LOCUS
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION AE000205 U00096
VERSION AE000205.1 GI:1787265
SOURCE
ORGANISM Escherichia coli K12
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
AUTHORS Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
Gregor, J., Davis, N.W., Klinkpatrick, H.A., Goeden, M.A., Rose, D.J.,
Mau, B. and Shao, Y.
The complete genome sequence of Escherichia coli K-12
Science 277 (5331), 1453-1474 (1997)
TITLE
JOURNAL 9742617
PUBMED 9278503
REFERENCE
AUTHORS Blattner, F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
3 (bases 1 to 10346)
Blattner, F.R.
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
4 (bases 1 to 10346)
Plunkett, G. III.
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GenMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
30332 (e-mail: mark@amber.gatech.edu). Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
its annotations are periodically updated; this is version M54. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
source
1..10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"

repeat_region
promoter
promoter
gene
CDS

protein_bind
protein_bind
protein_bind
promoter
gene
CDS

/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1..48)
/feature="IS3"
<1..5
/feature="factor Sigma70; predicted +1 start at 1094689"
5..32
/feature="factor Sigma70; predicted +1 start at 1094716"
69..392
/feature="b1028"
69..392
/feature="b1028"
/feature="orf; unknown"
/feature="orf; This 107 aa ORF is 27 pct identical (0 gaps)
to 58 residues of an approx. 1888 aa protein FAS2_CANAL
SW: P43098"
/codon_start=1
/translation="MEDGVQAMRDYLAQGLDASPERHQLVNNYAKSEVAPSLIKENLS
LHRTVVKWTSLEDTFLAMPVPVAFLEISNKPYLGNLNDPAGVDQQRVHCRKARSD
AVVFK"
275..299
/feature="b1028"
/feature="factor Sigma32; predicted +1 start at 1094983"
389..1375
/feature="ycdY"
/feature="synonym: b1029"
389..1375
/feature="ycdY"
/feature="orf; unknown"
/feature="orf; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCB
SW: P47562"
/codon_start=1
/translation="MPTDLYTRPQDRNLIYVAILGLLIGFYMNAGFTPESDI
LITLSMNGSLVNLFPDLKAVYACVKNIDPSFWFKKQKHLEFLTOYVAGFS
SLIRYVLVLCISTVNLITLISPLNSDSFTGEGOMLTPKSIALLILICGVATINLF
FTRFSKRAVFLGRLFLQELIDFESSENLSTPEPAKPIMLRFLILVLEVMITLSVL
ALTVEMRIWFEAVYFLTCVPLCIYFFYCRFLMHNDFMMACDMYFRMGHNK"
1253..1280
/feature="ycdY"
/feature="central position to predicted promoter: -104"
/feature="bound moiety="Deor predicted site"
1253..1280
/feature="ycdY"
/feature="No predicted promoter"
/feature="bound moiety="Deor predicted site"
1337..1365
/feature="ycdY"
/feature="factor Sigma70; predicted +1 start at 1096049"
1494..1745
/feature="b1030"
1494..1745
/feature="b1030"
/feature="orf; unknown"
/feature="orf; This 83 aa ORF is 25 pct identical (7 gaps)
to 61 residues of an approx. 616 aa protein ALBU_RABIT SW:
P49065"
/translation="MGGRRCVQYTKKKPVSYLFRYGGEGIDLSRSPCCQPAHC
/codon_start=1
/translation="b11"
/feature="orf; hypothetical protein"
/feature="protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGRRCVQYTKKKPVSYLFRYGGEGIDLSRSPCCQPAHC

```


Itch, T., Kanai, K., Kasai, H., Kashimoto, K., Kim, S., Kimura, S., Kitagawa, M., Kitakawa, M., Makino, K., Masuda, S., Miki, T., Mizobuchi, K., Mori, H., Motomura, K., Nakamura, Y., Nashimoto, H., Nishio, Y., Oshima, T., Saito, N., Sampel, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C., Yamamoto, Y., and Yano, M.
 Headed by:
 Name: Takashi Horiuchi
 Address: National Institute of Basic Biology, Okazaki, 444, Japan
 E-mail: kihori@nibb.ac.jp
 Information operator:
 Name: Hirokazu Mori
 Address: NARA Institute of Science and Technology,
 Ikoma, 630-01, Japan
 E-mail: hmori@itc.aist-nara.ac.jp
 URL:
 The Japan E. coli genome database
 http://baw3.aist-nara.ac.jp.
 Location/Qualifiers
 1..15047
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /db_xref="taxon:83333"
 /map="23.7-24.0 min"
 /clone="Kohara clone #231"
 /note="Nucleotide position 1097101-1112147 from the initiation site of Thra (0 min.). This clone is from Kohara lambda minist library."
 <1..276
 /note="unnamed protein product; ORF_ID:0230#5"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA35822.1"
 /db_xref="GI:4062605"
 /translation="DLFFSENNALSTPFAKPLMLRFLILYEVMMITLVSVALIVEWRIMFEAYFLCLVPCLIYFYFCRFLMNDPMACMDYFRMGHFNK"1152..2090
 /note="ORF_ID:0231#1
 similar to SwissProt Accession Number P45637"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical 33.0 kd protein in prob-proA intergenic region."
 /protein_id="BAA35823.1"
 /db_xref="GI:4062606"
 /translation="MDIIFHPTPTDQWMTLEAKLPAQARVRANKGNDSDYALYVHPVEMLAGRLKAVFALGCVSIIISKQAHPEMLNPVPLFLEDTGMGOMKEYAVSOVLHWRRRDPDYRIQNSSHWQPLPEYHREDFTIGLAGVLGSKVAOSIQMRFPLRCMSPTKSPKMPQVQSPAGREELSAFLSQCRVLINLPTPTETVGINOOLLEKLPDCAVLTNLRAGVHVVEDLLALDGGKVGKMLDVFNREPLPPESPIMQHPRVTITPHVAAITRPAVAEYISRTIAOLEKGERVCGQVDRAGY"2145..2882
 /note="unnamed protein product; ORF_ID:0231#2"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA35824.1"
 /db_xref="GI:4062607"
 /translation="MYVDLHMTVASTHAYSTLSDYIAOKOKGIXLPAITDGPDMEDAPHHNFTNMRIPWVVDGVLIRGIEANIKNDGETDCGRFSDILILIGFHSFHSRSGSDNCRVAAAVRACGAVALGSDSHIAFTWGEFECKLIDAVDPPERILNVSPRLNLPFESRGMAPIAEFADL"2906..3460
 /note="ORF_ID:0231#3
 similar to PIR Accession Number F64035"
 /codon_start=1
 /transl_table=11
 /product="Hypothetical protein HI1543"
 /protein_id="BAA35825.1"
 /db_xref="GI:4062608"
 /translation="MNEPSSILCRVIGSLYTYQPDPLLVPLFTLIRGKLIANMPLEODLLRLRLQKSCDMQVADYNALFTIGDECAVPPYRSAMVEGATEAEVRAFLSRGNPL

CDS

gene

CDS

gene

CDS

gene

CDS

CDS

gene

CDS

ADTPADHICTLLLAASWIEDSTEDSEAELETIFSEYLLPMWCAFLGKVEAHATTPFW
 RMAULITDAISAMWDELEBDSER"3514..4053
 /note="unnamed protein product; ORF_ID:0231#4"
 /codon_start=1
 /transl_table=11
 /protein_id="BAA35826.1"
 /db_xref="GI:4062609"
 /translation="MCSIFFMAAFSAIMGNMNLISAITTGILSGIMGVAVSICLIVAFMLCTIAKQKLLSPVPGTFIGACATPAGCGDWKLVPLSLALGIIICYAMKNSGLWLAARSATYAREBEIKKA"4117..4950
 /gene="cs9G"
 /complement (4117..4950)
 /note="ORF_ID:0231#5
 similar to SwissProt Accession Number P52103"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9G precursor."
 /protein_id="BAA35827.1"
 /db_xref="GI:4062610"
 /translation="WQRLFLVAVMLSGCLTAPPKEARPTLMPRAQSYKDLTHLPA
 NLRKTFVSYTINQDETQFKPYPAISNFSIVAPQSAATMLVTALKDSRWPTLERQILO
 ICADYQYLDQIAVNLRVNVSSTGSIISVNTKTLISYVQGVRFIDYQKLEGE
 VGYTNSPEVMLCLMSALETGVIFPLINDIGRLMDLONKAEKRONDLIVYRHMSVPE
 S"4977..5393
 /gene="cs9F"
 /complement (4977..5393)
 /note="ORF_ID:0231#6
 similar to SwissProt Accession Number P52104"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9F precursor."
 /protein_id="BAA35828.1"
 /db_xref="GI:4062611"
 /translation="MKRVKAVVLLMLISPLSWAGTMTFOFRRNPFGNPNNGCALINS
 AQQNSYKQPSYNDPGETPSALDNFTQAIQSOLIGGLSNINTGKRGKRVATNDYIV
 DIANRGGQQLQVLTDRKGTQSTIQVSGIQNNSTDF"5418..5807
 /gene="cs9E"
 /complement (5418..5807)
 /note="ORF_ID:0231#7
 similar to SwissProt Accession Number P52105"
 /codon_start=1
 /transl_table=11
 /product="Curli production assembly/transport component
 cs9E precursor."
 /protein_id="BAA35829.1"
 /db_xref="GI:4062612"
 /translation="MKRILRWIVAAEFLFAAGNLAHVEVEPGLLTDHVVSSIGHDY
 RASFQMSDITGLTITNERPSARWGSMTITIVNQVIFQTFPLFKRPFEXTVVPAL
 IOTEBALNRQRIINQALSTGDLAHEF"5812..6462
 /gene="cs9D"
 /complement (5812..6462)
 /note="ORF_ID:0231#8
 similar to SwissProt Accession Number P52106"
 /codon_start=1
 /transl_table=11
 /product="Probable cs9AB operon transcriptional regulatory
 protein."
 /protein_id="BAA35830.1"
 /db_xref="GI:4062613"
 /translation="MNEVHSHIGHTLLITKSSIQATVALLQHLKOSIAITGKCLHNIO

THIS PAGE BLANK (USPTO)

CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Bacteriophage* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
 Best Local Similarity 100.0%; Pred. No. 2.9e-137;
 Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
 QY 61 GGGCTGTTCCCAATGGGGCGGCGGATATCAATACCGCGGCAATAGTTCCGCG 120
 DB 61 GGGCTGTTCCCAATGGGGCGGCGGATATCAATACCGCGGCAATAGTTCCGCG 120
 QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCTGTGCA 180
 DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCTGTGCA 180
 QY 181 AGCGATCCCGTAAATCTGAAAGCAATCCAGAGGGGTTATGTAAGCGCGCAT 240
 DB 181 AGCGATCCCGTAAATCTGAAAGCAATCCAGAGGGGTTATGTAAGCGCGCAT 240
 QY 241 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 DB 241 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 QY 301 GCGACATGACCACTGTAAGCGTAAAACTCCGATTTACTGTCGGCCAAATCGCGGT 360
 DB 301 GCGACATGACCACTGTAAGCGTAAAACTCCGATTTACTGTCGGCCAAATCGCGGT 360
 QY 361 AATTAAGCGCGCTGTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
 DB 361 AATTAAGCGCGCTGTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
 QY 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456
 DB 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456

RESULT 2
 AA087467
 ID AA087467 standard; DNA; 456 BP.
 AC AA087467;
 XX
 DT 25-MAR-2003 (revised)
 OS 26-JUN-1995 (first entry)
 DE Agfa sequence.
 XX
 KW *Salmonella*; Agfa; vaccine; genetic immunization; ds.
 OS *Salmonella*.
 XX

PH Key Location/Qualifiers
 FT CDS 1..454
 FT /tag= a
 FT /note= "Agfa"
 XX
 XX MO9425598-A2.
 XX
 XX 10-NOV-1994.
 XX
 XX 26-APR-1994; 94WO-IB000207.
 XX
 XX 26-APR-1993; 93US-00054452.
 XX
 XX (UWV1-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX (KING) KING J.
 XX
 XX Kay WW, Collinson SK, Clouthier SC, Doran JL,
 XX WPI: 1994-358275/44.
 XX P-PSDB; AAR74625.
 XX
 XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 XX strains, vector constructs, or compms. contg. fimbrial type proteins.
 XX Disclosure; Fig 7B; 95pp; English.
 XX
 XX The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated
 XX protein are used in genetic immunization and vaccine compositions,
 XX respectively, to elicit an immune response to *Salmonella* in animals (e.g.
 XX food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
 XX field.)
 XX
 XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.8%; Score 404.8; DB 2; Length 456;
 Best Local Similarity 93.0%; Pred. No. 1.1e-120;
 Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGAGTGTCTGGCT 60
 QY 61 GGGCTGTTCCCAATGGGGCGGCGGATATCAATACCGCGGCAATAGTTCCGCG 120
 DB 61 GGGCTGTTCCCAATGGGGCGGCGGATATCAATACCGCGGCAATAGTTCCGCG 120
 QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCTGTGCA 180
 DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCTGTGCA 180
 QY 181 AGCGATCCCGTAAATCTGAAAGCAATCCAGAGCGGTTATGTAAGCGCGCAT 240
 DB 181 AGCGATCCCGTAAATCTGAAAGCAATCCAGAGCGGTTATGTAAGCGCGCAT 240
 QY 241 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 DB 241 GTAGGCGAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 QY 301 GCGACATGACCACTGTAAGCGTAAAACTCCGATTTACTGTCGGCCAAATCGCGGT 360
 DB 301 GCGACATGACCACTGTAAGCGTAAAACTCCGATTTACTGTCGGCCAAATCGCGGT 360
 QY 361 AATTAAGCGCGCTGTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
 DB 361 AATTAAGCGCGCTGTGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
 QY 421 GCACATGCAAAACGCGACGCTAACCAAGTATTA 456
 DB 421 GGTGTTGCAAAACGCGACGCTAACCAAGTATTA 456

RESULT 3
 AAT74142


```

ID  AAT74142 standard; DNA; 456 BP.
XX
AC  AAT74142;
XX
DT  25-MAR-2003 (revised)
DT  29-SEP-1997 (first entry)
XX
DE  Salmonella enteritidis 27655-3b agfA gene.
XX
KM  Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
OS  Salmonella enteritidis.
XX
Key  Location/Qualifiers
FH  1..456
FT  /*tag= a
FT  /label= agfA_gene_fragment
FT  /transl_except= (pos:367..369,aa:Pro)
XX
PN  US563617-A.
XX
PD  03-JUN-1997.
XX
PF  26-APR-1994; 94US-00233788.
XX
PR  26-APR-1993; 93US-00054452.
XX
PA  (UTVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
PI  Collinson SK, Kay WW, Doran JL;
XX
DR  WPI: 1997-309886/28.
DR  P-PSDB; AAM23570.
XX
PT  Isolated Salmonella gene agfA - used for diagnosis of Salmonella or
XX  enteropathogenic bacteria of the Enterobacteria family.
XX
PS  Claim 1; Col 19-112; 85pp; English.
XX
CC  The present sequence represents an isolated agfA gene derived from
CC  Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC  diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC  family Enterobacteria. It can also be used to provide proteins and
CC  antibodies which can be used for assays. The nucleic acid sequence can be
CC  used to provide probes or primers which can specifically hybridise to
CC  nucleic acid molecules from greater than 9% of Salmonella strains that
CC  are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC  from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC  correct PF field.)
XX
SQ  Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.1e-120;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

```

Db  241 GTAGCCAGGGTGGGATATATGACTTGAACCTGACTGAGATGGTTTCAGAAATAT 300
Qy  301 GCCACATCGACGATGGAACGCTAAAACTCCGATATATTCTGCGCCAAATACGGCGT 360
Db  301 GCCACATCGACGATGGAACGCTAAAACTCCGATATATTCTGCGCCAAATACGGCGT 360
Qy  361 AATTAACCGCGGCTGTTAATATATGATCAGCTGTACCCGTGTGTTGTTACCATGAATG 420
Db  361 AATTAACCGCGGCTGTTAATATATGATCAGCTGTACCCGTGTGTTGTTACCATGAATG 420
Qy  421 GCACATGCAAAACAACGCGCTAACGATTTAA 456
Db  421 GGTTTGGCAACAAACGCGCTAACGATTTAA 456

RESULT 4
AAC64617
ID  AAC64617 standard; DNA; 456 BP.
XX
AC  AAC64617;
XX
DT  26-FEB-2001 (first entry)
XX
DE  Salmonella enteritidis AgfA DNA sequence SEQ ID NO:1.
XX
KM  Salmonella; agfA; chromosomal gene replacement; fimbria; epitope;
XX  vaccine; immune response; immunogen; ds.
XX
OS  Salmonella enteritidis.
XX
PN  WO200060102-A2.
XX
PD  12-OCT-2000.
XX
PF  05-APR-2000; 2000WO-CA000356.
XX
PR  05-APR-1999; 99US-0127888P.
XX
PA  (UTVI-) UNIV VICTORIA.
XX
PI  White AP, Doran JL, Collinson SK, Kay WW;
XX
DR  WPI: 2000-672631/65.
DR  P-PSDB; AAB36341.
XX
PT  Recombinant agfA gene having a segment replaced by a foreign DNA sequence
XX  which encodes foreign epitope or antigen, expresses recombinant AgfA
XX  protein useful for eliciting immune response in animal.
XX
PS  Disclosure; Page 134; 139pp; English.
XX
CC  The present invention describes a recombinant agfA gene (I) where a
CC  segment of the gene has been replaced by a segment of a foreign DNA
CC  sequence which encodes a foreign epitope or antigen. Also described are:
CC  (1) use of thin aggregative fimbriae (SfP17/TAF) nucleation dependent
CC  assembly system of strains of Salmonella, Escherichia coli and
CC  Enterobacteriaceae for the production of fimbriae comprising: (2)
CC  AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (3)
CC  directing recombination of a recombinant gene into the chromosome of the
CC  homologous species; (4) directing recombination of a recombinant gene
CC  back into the chromosome of the homologous species, replacing the native
CC  copy of that gene; and (4) eliciting an immune response in an animal,
CC  comprising separating an amino acid polymer comprising a recombinant
CC  protein containing a replacement segment or segments of foreign amino
CC  acid sequence or sequences grown on a Salmonella, E. coli or
CC  Enterobacteriaceae host cell, from the host cell and introducing the
CC  polymer into the animal in conjunction with a carrier or diluent. (I) is
CC  useful for the expression of recombinant AgfA protein which is useful for
CC  eliciting an immune response in an animal. In a fimbrial presentation
CC  system the heterologous antigens are presented in high numbers (up to
CC  500,000 copies/cell), the hybrid fimbria protein possesses both the
CC  immunogenicity and adhesion properties relevant for an efficient live

```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP, 117 A, 112 C, 122 G, 105 T, 0 U, 0 Other;

Query Match Best Local Similarity 88.8%; Score 404.8; DB 3; Length 456;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTAAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTAAGTTTCTGCGAGTCTTGCT 60
QY 61 GGGCTGCTTCCACATGGGGCGCGCGGCTTAATCAACGCGCGGCAATAGTTCCGGC 120
DB 61 GGGCTGCTTCCACATGGGGCGCGCGGCTTAATCAACGCGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTTATCATAGTTCGCTTCCGCTTCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTTATCATAGTTCGCTTCCGCTTCTGCTGCA 180
QY 181 AGCGATGCGCCGTAATCTGAAACGACCATTAACGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCGCCGTAATCTGAAACGACCATTAACGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATATATGTAATGTAAGTCAAGTCAAGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATATGTAATGTAAGTCAAGTCAAGATGTTTCAAAATAT 300
QY 301 GCCACATTCGACCAATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGGCT 360
DB 301 GCCACATTCGACCAATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGGCT 360
QY 361 AATAACCGCGCGCTGTTAATTATGATGATGCTGTTAACCCTGTTTACCAATGAATG 420
DB 361 AATAACCGCGCGCTGTTAATTATGATGATGCTGTTAACCCTGTTTACCAATGAATG 420
QY 421 GCACATGCAACCAACGCGCGCTTAACGAGCATTTATTA 456
DB 421 GCACATGCAACCAACGCGCGCTTAACGAGCATTTATTA 456

```

RESULT 5
 AAC64623
 ID AAC64623 standard; DNA, 456 BP.

```

XX AC AAC64623;
XX DT 26-FEB-2001 (first entry)
XX DE AGFA::PT3#2 DNA sequence SEQ ID NO:13.
XX KM Salmomella; agfa: chromosomal gene replacement; fimbria; epitope;
XX KW vaccine; immune response; immunogen; ds.
XX OS Salmomella enteritidis.
XX OS Escherichia coli.
XX OS Synthetic.
XX PN WO200060102-A2.
XX PD 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UUVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW,

```

XX WPI: 2000-672631/65.
 DR P-PSDB: AAB35347.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (It is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell) the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP, 118 A, 109 C, 121 G, 108 T, 0 U, 0 Other;

Query Match Best Local Similarity 82.5%; Score 376; DB 3; Length 456;

Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTAAGTTTCTGCGAGTCTTGCT 60
DB 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATGTAAGTTTCTGCGAGTCTTGCT 60
QY 61 GGGCTGCTTCCACATGGGGCGCGCGGCTTAATCAACGCGCGGCAATAGTTCCGGC 120
DB 61 GGGCTGCTTCCACATGGGGCGCGCGGCTTAATCAACGCGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTTATCATAGTTCGCTTCCGCTTCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTTATCATAGTTCGCTTCCGCTTCTGCTGCA 180
QY 181 AGCGATGCGCCGTAATCTGAAACGACCATTAACGAGCGGTTATGTAAGCGCGCAT 240
DB 181 AGCGATGCGCCGTAATCTGAAACGACCATTAACGAGCGGTTATGTAAGCGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATATATGTAATGTAAGTCAAGTCAAGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATATGTAATGTAAGTCAAGTCAAGATGTTTCAAAATAT 300
QY 301 GCCACATTCGACCAATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGGCT 360
DB 301 GCCACATTCGACCAATGGAACGCTTAAACTCCGATTTACTGTCGCGCAATAGCGGCT 360
QY 361 AATAACCGCGCGCTGTTAATTATGATGATGCTGTTAACCCTGTTTACCAATGAATG 420
DB 361 CTGGTACCGGCTGTTTATGATGATGCTGTTAACCCTGTTTACCAATGAATG 420
QY 421 GCACATGCAACCAACGCGCGCTTAACGAGCATTTATTA 456

```

DB 421 GGTTTGGCAACGACCAGCGCTAACAGTAATTAA 456

RESULT 6
ID AAC64626 standard; DNA; 456 BP.
XX AAC64626;
XX AC
XX XX
D7 26-FEB-2001 (first entry)
DE Agfa::PT#5 DNA sequence SEQ ID NO:19.
XX
XX Salmomella: agfa: chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
OS Salmomella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX MO200060102-A2.
PN 12-OCT-2000.
PD 05-APR-2000; 2000MO-CA000356.
XX PF
PR 05-APR-1999; 99US-0127888P.
XX (UTVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX WPI: 2000-672631/65.
DR P-FSDB: AAB36350.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure: Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended assembly system of strains of Salmomella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmomella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

QY	1	TTGAAACTTTTAAAGTGGCAGCATTCGGAGCAATCGATGTTCTTGCGAGTGTCTTGCGCT	60
Db	1	ATGAAACTTTTAAAGTGGCAGCATTCGGAGCAATCGATGTTCTTGCGAGTGTCTTGCGCT	60
QY	61	GCGCTGTTCCCAATGGGGCGCGCGGTAAATCATTAACGGCGCGCGGCATAGTTCGCGC	120
Db	61	GCGCTGTTCCCAATGGGGCGCGCGGTAAATCATTAACGGCGCGCGGCATAGTTCGCGC	120
QY	121	CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTACGCTGCGCTTGCTTCGCA	180
Db	121	CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTACGCTGCGCTTGCTTCGCA	180
QY	181	AGCGATGCCCGTAAATCTGAAACGACATTTACCCGAGCGGTTATGTAACGGCGCGCAT	240
Db	181	AGCGATGCCCGTAAATCTGATGATGATGACTGCTGTTACCCGCTGTTGTAACCATGAAATGGCACAT	240
QY	241	GTAGGCCAGGGTGGCGATPATATGATCTATTTGAATGACTCAGAAATGGTTTCAGAAATTAAT	300
Db	241	GCAGGCCAGGGGTCGGATPATATGATCTATTTGAATGACTCAGAAATGGTTTCAGAAATTAAT	300
QY	301	GCACACATCGACCAATGGAAACGCTAAAAATCTCCGATATTAATCTGGGCCCAATAGCGCGAT	360
Db	301	GCACACATCGACCAATGGAAACGCTAAAAATCTCCGATATTAATCTGGGCCCAATAGCGCGAT	360
QY	361	AATAACGCCGCGCTGTTAATTATGATCAGCTGCTTACCCGCTGTTGTTACCATGATAATG	420
Db	361	AATAACGCCGCGCTGTTAATTATCAGACCGCATCTGATTCAGCGTAAATGGTGTGATGAGTT	420
QY	421	GCACATGCAAAACACGCCACGGCTTAACCATGATTTAA	456
Db	421	GGTTTGGCAACACGCCACGGCTTAACCATGATTTAA	456

RESULT 7
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
XX AAC64625;
AC
XX
DT 26-FEB-2001 (first entry)
DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PV WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WJ;
XX
DR WPI; 2000-672631/65.
DR P-PsDB; AAB36349.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
P8 Disclosure; Page 136; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (I) where a

Db	121	CCGGA	CTCAA	CGTTG	AGCATTTAT	CAGTACGGTTCGGCTAA	CGCTGGCTGCTGCA	180
Qy	181	AGCGAT	CCCCG	CTTAAATCT	TGAACGACCATTTAC	CCAGACGGGTATATG	GTAAACGGCCCGAT	240
Db	181	AGCGAT	CCCCG	CTTAAATCT	TGAACGACCATTTAC	CCAGACGGGTATATG	GTAAACGGCCCGAT	240
Qy	241	GTAGC	CCAGG	GTGCGGATAT	AGTACTATT	TGAACTACTCAGAA	TGTTTCAAAATTAAT	300
Db	241	GTAGC	CCAGG	GTGCGGATAT	AGTACTATT	TGAACTACTCAGAA	TGTTTCAAAATTAAT	300
Qy	301	GCCAC	CATCTG	ACCACTG	GAAGGTAAAAATCT	CCGATATTA	CTGTCCGCCCAATTCGGCCGT	360
Db	301	GCCAC	CATCTG	ACCACTG	GAAGGTAAAAATCT	CCGATATTA	CTGTCCGCCCAATTCGGCCGT	360
Qy	361	AATAA	CGCCG	CGCTGTGTTAT	TATATGATCAGCTG	GTATACCGGTGTTT	TACCACTGAATG	420
Db	361	CATGA	AATG	CACATG	CAAAATCAGACCG	GTATTCACAGGT	TAAATGTCGCTCAGGTT	420
Qy	421	GCACAT	GCAAA	CAAGCC	ACGGCTAAC	CAGTATTA	456	
Db	421	GCTTTT	GCAAC	AGCCAC	GGCTAA	CCAGTATTA	456	

RESULT 9

ID AAC64629 standard; DNA; 456 BP.

AC AAC64629;

DT 26-FEB-2001 (first entry)

AgfA::PT3#8 DNA sequence SEQ ID NO:25.

KW *Salmonella*; *agfA*; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.

OS *Salmonella enteritidis*

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI -) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;
.....

DR WPI; 2000-672631/65.

XX

PT which encodes foreign epitope or antigen, expresses recombinant AgfA

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

The present invention describes recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TA9) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa

protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrinrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrinrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrinriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match	77.5%	Score 353.6	DB 3	Length 456

Matches 392; Conservative 0; Mismatches 64;

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGCAATCGTAGTTTCTGGCAGTGCTCTGGCT 60

Oy	1	ATGAAACCTTTAAAAAGTGGACGACATTCGACGAATCGTAGTTTCTGGACGTCTCTGGCT	60
Db	1	ATGAAACCTTTAAAAAGTGGACGACATTCGACGAATCGTAGTTTCTGGACGTCTCTGGCT	60
Oy	61	GGCGTCGTTCCACAATGGGGCGGCGCGGTATCATTAACGGCGGCGGCAGATAGTTCCGGC	120
Db	61	GGCGTCGTTCCACAATGGGGCGGCGCGGTATCATTAACGGCGGCGGCAGATAGTTCCGGC	120
Oy	121	CCGGAATCAAGTTGAGAGATTATCAGTACGGTTCCGGTAACGTCGGCTTCTCTGCA	180
Db	121	CCGGAATCAAGTTGAGAGATTATCAGTACGGTTCCGGTAACGTCGGCTTATATATCAG	180
Oy	181	AGCGATGCCCGTAAATCTGAAACGACCATTAACCCAGACGGTTATGTGTAAACGGCCGAT	240
Db	181	CTGGTTACCGCGTGTGTATTACCAATGAATATGGACATGACAGGTTATGTGTAAACGGCCGAT	240
Oy	241	GTAGGCCAGGGGTGGGATATATGATCTATTGACTGACATAGGTTTCAAAATAT	300
Db	241	GTAGGCCAGGGGTGGGATATATGATCTATTGACTGACATAGGTTTCAAAATAT	300
Oy	301	GCCACCAATCGACGAGTGGACGCTAATAAACTCCGATATTACTGTGGGCAATATACGGCGAT	360
Db	301	GCCACCAATCGACGAGTGGACGCTAATAAACTCCGATATTACTGTGGGCAATATACGGCGAT	360
Oy	361	AATAACGCCGCGTGTATTATATATGATCAGCTGTTTACCCGTGTGTATACCATGAAATG	420
Db	361	AATAACGCCGCGTGTATTATCAGACCGCATCTGATTTCCAGCGTATATGATGCTCAGGTT	420
Oy	421	GCACATGCAAAACAACGCCACGGCTAATCCAGATTAA	456
Db	421	GGTTTGGCAACAACGCCACGGCTAATCCAGATTAA	456

RESULT 10

ID AAC64624 Standard; DNA; 456 BP.

AC AAC64624;

DT 26-FEB-2001 (first entry)

DE AgFA::PT3#3 DNA sequence SEQ ID NO:15.

KW *Salmonella*; agfA; chromosomal gene replacement; fimbria; epitope;

XX

OS Escherichia coli.

XX
XX

XX

PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 XX (UUVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 DR WPI; 2000-672631/65.
 XX P-PSDB; AAB36354.
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TFP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SQ Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 1.5e-103;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGGCGCATTCGAGCAATCGTATGTTTGGCAGTCTGGCT 60
 DB 1 ATGAACTTTAAAGTGGCGCATTCGAGCAATCGTATGTTTGGCAGTCTGGCT 60
 QY 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATGTTCCGGC 120
 DB 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATGTTCCGGC 120
 QY 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATGTTCCGGC 120
 DB 61 GGGCTGCTTCAACATGGGCGCGCGGGAATCAAAAGGCGCGCAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAACCGCTTCTGCTCAA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAACCGCTTCTGCTCAA 180
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAACCGCTTCTGCTCAA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGCTAACCGCTTCTGCTCAA 180
 QY 181 AGGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTTGTGTAACGCGCGCAT 240
 DB 181 AGGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTTGTGTAACGCGCGCAT 240
 QY 181 AGGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTTGTGTAACGCGCGCAT 240
 DB 181 AGGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTTGTGTAACGCGCGCAT 240
 QY 241 GTAGGCGGAGGTGGGATTAATGTTCTATTTGAATGTAAGTCAAGATGGTTTCAGAAATAT 300
 DB 241 GTAGGCGGAGGTGGGATTAATGTTCTATTTGAATGTAAGTCAAGATGGTTTCAGAAATAT 300
 QY 301 GCACACATGACCAAGTGGAGCGTAAACCGCATTTACTGCGGCCAATACGCGCGT 360

DB 301 |||||
 QY 361 AATAAGCGCGGCTGTTAATTAATGATCAGTGGTTACCGTGTGTTAACCAGTAATG 360
 DB 361 AATAAGCGCGGCTGTTAATTAATGATCAGTGGTTACCGTGTGTTAACCAGTAATG 420
 QY 421 GCACATGCAACACAGCGGCTAACCAGTATTA 456
 DB 421 GCTTTGCAACACAGCGGCTAACCAGTATTA 456

RESULT 11
 ID AAC64630 standard; DNA; 456 BP.
 XX AAC64630;
 AC 26-FEB-2001 (first entry)
 XX 26-FEB-2001 (first entry)
 DT Agfa::PT39 DNA sequence SEQ ID NO:27.
 XX Agfa::PT39 DNA sequence SEQ ID NO:27.
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000WO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 XX (UUVI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay WW;
 DR WPI; 2000-672631/65.
 DR P-PSDB; AAB36354.
 XX Disclosure; Page 138; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TFP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 4.8e-103;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGGCAGTCTGGCT 60
QY 61 GGCTTCGTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGCTTCGTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAACGACCATTAACGAGCGGTTATGTTACGGCCGAT 240
DB 181 AGCGATGCCGTTAAATCTGAACGACCATTAACGAGCGGTTATGTTACGGCCGAT 240
QY 241 GTAGCCGAGGTCGGCATTAATGACTTAATGACTGACTGCAATGGTTTGAATAT 300
DB 241 GTAGCCGAGGTCGGCATTAATGACTTAATGACTGACTGCAATGGTTTGAATAT 300
QY 301 GCCGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 360
DB 301 GCCGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 360
QY 361 AATTAAGCCGCGGTGTTAATTAATGATCAAGTGTGTTACCGCTGTTTGAATGAATG 420
DB 361 AATTAAGCCGCGGTGTTAATTAATGATCAAGTGTGTTACCGCTGTTTGAATGAATG 420
QY 421 GCACATGCAAAACAGCCGACGCTTAACGATTAATTA 456
DB 421 GCATTTGGCAACAGCCGACGCTTAACGATTAATTA 456

RESULT 12
AAC64627
ID AAC64627 standard; DNA; 456 BP.
AC AAC64627;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#6 DNA sequence SEQ ID NO:21.
XX
XX Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO20060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UVT-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFF1/TFP) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SO Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 4.8e-103;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGAGCATTCGACCAATCGTAGTTCTGGCAGTCTGGCT 60
QY 61 GGCTTCGTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GGCTTCGTTCCAAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAACGACCATTAACGAGCGGTTATGTTACGGCCGAT 240
DB 181 AGCGATGCCGTTAAATCTGAACGACCATTAACGAGCGGTTATGTTACGGCCGAT 240
QY 241 GTAGCCGAGGTCGGCATTAATGACTTAATGACTGACTGCAATGGTTTGAATAT 300
DB 241 GTAGCCGAGGTCGGCATTAATGACTTAATGACTGACTGCAATGGTTTGAATAT 300
QY 301 GCCGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 360
DB 301 GCCGACTCAACGTTGAGCATTTATCAATGACGTTCCGTTACGCTGCTCTGCA 360
QY 361 AATTAAGCCGCGGTGTTAATTAATGATCAAGTGTGTTACCGCTGTTTGAATGAATG 420
DB 361 AATTAAGCCGCGGTGTTAATTAATGATCAAGTGTGTTACCGCTGTTTGAATGAATG 420
QY 421 GCACATGCAAAACAGCCGACGCTTAACGATTAATTA 456
DB 421 GCATTTGGCAACAGCCGACGCTTAACGATTAATTA 456

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT#10 DNA sequence SEQ ID NO:29.
XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
PN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
DR P-PSDB; AAB36355.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139pp; English.
XX
CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising Agfa
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 76.1%; Score 347.2; DB 3; Length 456;
Best Local Similarity 85.1%; Pred. No. 5.3e-102;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
OY 1 ATGAACCTTTAAAGTGCGACGATTCGACGCAATCGATGTTCTGCGAGTGTCTGGCT 60
XX

Db 1 ATGAACCTTTAAAGTGCGACGATTCGACGCAATCGATGTTCTGCGAGTGTCTGGCT 60
OY 61 GCGCGTCTTCCACATATGCGCGCGCGGATATCATATACGCGCGGCAATATCTCCGCG 120
Db 61 GCGCGTCTTCCACATATGCGCGCGCGGATATCATATACGCGCGGCAATATCTCCGCG 120
OY 121 CCGGACTCAAGCTTGAGCATTTATGAGTATGAGTTCCTGCTGAGTTCGCTGCA 180
Db 121 CCGGACTCAAGCTTGAGCATTTATGAGTATGAGTTCCTGCTGAGTTCGCTGCA 180
OY 181 AGCGATGCGCGTAAATCTGAAACGACCATTTACCGAGCGGTTATGTAACGCGCGCAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACCATTTACCGAGCGGTTATGTAACGCGCGCAT 240
OY 241 GTAGGCGAGGATCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGGCGAGGATCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
OY 301 GCCACCATGACGACGATGGAACGCTTAAATCCGATATTACTGTCGCGCATATAGCGCGGT 360
Db 301 GCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
OY 361 AATTAACGCGCGCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 361 AATTAACGCGCGCTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
OY 421 GCAACATGCAACACGACGCGCTTAACCATTAATTA 456
Db 421 GCTTTTGCAACACGACGCGCTTAACCATTAATTA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
AC AAQ73066;
XX
DT 27-AUG-2003 (revised)
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX Salmonella; Agfa; vaccine; genetic immunization; ds.
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1..359
FT /tag= a
FT /note= "Agfa"
FT misc_feature 37..60
FT /tag= d
FT /note= "TAF5 primer (pair with TAF6)"
FT misc_feature 52..69
FT /tag= b
FT /note= "TAF3 primer (pair with TAF4)"
FT misc_feature complement(103..129)
FT /tag= e
FT /note= "TAF6 primer (pair with TAF5)"
FT complement(292..402)
FT /tag= c
FT /note= "TAF4 primer (pair with TAF3)"
XX
PN WO9425598-A2.
XX
XX 10-NOV-1994.
PD 26-APR-1994; 94WO-IB000207.
PR 26-APR-1993; 93US-00054452.
XX

Thu Mar 18 12:27:55 2004

us-09-543-407-11.png

Page 12

Job time : 246.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgcctaacagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCUTS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	404.8	88.8	456	1	US-08-233-788A-58
2	315.2	69.1	361	1	US-08-233-788A-56
3	34	7.5	1344	3	US-09-120-927-1
4	34	7.5	1344	4	US-09-431-614-5
5	33.4	7.3	7766	4	US-09-125-619-3
6	33.4	7.3	14770	4	US-09-220-132-30
7	32.2	7.1	1122	4	US-09-489-039A-3211
8	31.2	6.8	1491	6	5486473-3
9	31.2	6.8	2436	4	US-09-540-236-492
10	31.2	6.8	65792	4	US-09-596-002-31
11	30.8	6.8	2547	3	US-08-508-761B-1
12	30.6	6.7	4403765	3	US-09-103-840A-2
13	30.6	6.7	4411529	3	US-09-103-840A-1
14	30	6.6	1341	4	US-09-350-756-2
15	29.8	6.5	1008	4	US-09-252-991A-10904
16	29.8	6.5	1077	4	US-09-252-991A-10818
17	29.8	6.5	1521	4	US-09-328-352-2055
18	29.6	6.5	972	4	US-09-252-991A-10865
19	29.4	6.4	423	4	US-09-252-991A-10818
20	29.4	6.4	1252	4	US-09-561-756-29
21	29.4	6.4	1252	4	US-09-227-721-29
22	29.4	6.4	1252	4	US-09-954-697-39
23	29.4	6.4	1939	3	US-08-961-527-310
24	29.4	6.4	3494	4	US-09-135-802-200
25	29.4	6.4	3494	4	US-09-659-786-200
26	29.4	6.4	3494	4	US-09-023-655-1006
27	29.4	6.4	3660	4	US-09-253-991A-12269

28	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appli
29	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appli
30	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appli
31	29	6.4	534	4	US-09-489-039A-2195	Sequence 2195, Ap
32	29	6.4	1200	4	US-09-134-000C-1704	Sequence 1704, Ap
33	29	6.4	1818	4	US-09-489-039A-2162	Sequence 2162, Ap
34	29	6.4	13794	4	US-08-956-171E-54	Sequence 54, Appl
35	28.8	6.3	654	4	US-09-252-991A-4506	Sequence 4506, Ap
36	28.8	6.3	720	4	US-09-328-352-354	Sequence 354, App
37	28.8	6.3	1662	1	US-08-565-386-2	Sequence 1, Appli
38	28.8	6.3	4529	1	US-08-565-386-1	Sequence 1, Appli
39	28.8	6.3	8906	4	US-09-027-169-5	Sequence 5, Appli
40	28.6	6.3	432	4	US-09-252-991A-4473	Sequence 4473, Ap
41	28.4	6.2	1026	3	US-07-751-891B-24	Sequence 24, Appl
42	28.4	6.2	4188	3	US-07-751-891B-2	Sequence 2, Appli
43	28.4	6.2	4242	3	US-07-705-490-2	Sequence 2, Appli
44	28.4	6.2	5222	3	US-07-751-891B-23	Sequence 23, Appl
45	28.2	6.2	357	4	US-09-489-039A-1118	Sequence 1118, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 3635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Cloutier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SBEANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 4.5e-111;
Matches 424; Conservative 0; Mismatches 32; Indels 0;

GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collins, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS OF SALMONELLA
NUMBER OF INVENTIONS: 61
TITLE OF INVENTIONS: OF SALMONELLA
CORRESPONDENCE ADDRESS:
ADDRESS:

STREET: Seed and Berry
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0,
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043,403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 652-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836
 INFORMATION FOR SEND ID NO: SEENBERY
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

US-08-233-788A-56

Oy 6 GTCTGTTCAACATGAGGCGCGCGCGGTATCATTAACGGCGCGCGCAATAGTTCCGGCCG 123
 Db 1 GTCTGATCAACAGTGGGCGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGCCG 60
 Oy 124 GACTCAACGTTTGGACATTTATCATGTAACGTTCCGCTACCGCTTCTCTGCAAGC 183
 Db 61 GACTCAACGTTTGGACATTTATCATGTAACGTTCCGCTACCGCTTCTCTGCAAGC 120
 Oy 184 GATGCCCGTAATCTGAAAACGACATTAACCGAGCGGTATGTTAAACGGCGCGCATGTA 243
 Db 121 GATGCCCGTAATCTGAAAACGACATTAACCGAGCGGTATGTTAAACGGCGCGCATGTA 180
 Oy 244 GGCACAGGTGCGGATATATGACTATTGAACTGACACAAATGGTTTCAGAAATATATGCC 303
 Db 181 GGCACAGGTGCGGATATATGACTATTGAACTGACACAAATGGTTTCAGAAATATATGCC 240
 Oy 304 ACCATCGACAGGTGGAACGCTAAAAAACCTCGATTTACTGTGCGCAATACGGCGGTAAT 363
 Db 241 ACCATCGACAGGTGGAACGCTAAAAAACCTCGATTTACTGTGCGCAATACGGCGGTAAT 300
 Oy 364 AACGGCGCGCGGTTAATTA 383
 Db 301 AACGGCGCGGTTAATTA 320

TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1344 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGGCTGCGCTGCTTCCCAATGGGGCGCGCGGTAATCATACGGCGGGCAAT 111
DB 209 GCTCCGGTTGGCGGATTAACCTGATGTGGCGATGACGTGCTTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGACCTCAAGCTGAGCATTTATCAGTACGGTTCGGCTAAGCGCTG 171
DB 149 AATTGACCATTTGCTCATGTGTTGCCGATCATGGGTTGCCCCCAACGAGAAATT 90

QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-431-614-5/c
; Sequence 5, Application US/09431614
; Patent No. 6624139
; GENERAL INFORMATION:
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Schading, Richard L.
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
; FILE REFERENCE: 21829/41 (EBC-003)
; CURRENT APPLICATION NUMBER: US/09/431,614
; CURRENT FILING DATE: 1999-11-02
; EARLIER APPLICATION NUMBER: 60/107,243
; EARLIER FILING DATE: 1998-11-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 1344
; TYPE: DNA
; ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGGCTGCGCTGCTTCCCAATGGGGCGCGCGGTAATCATACGGCGGGCAAT 111
DB 209 GCTCCGGTTGGCGGATTAACCTGATGTGGCGATGACGTGCTTAACAGTTCCGCCAAT 150
QY 112 AGTTCGGCGCGGACCTCAAGCTGAGCATTTATCAGTACGGTTCGGCTAAGCGCTG 171
DB 149 AATTGACCATTTGCTCATGTGTTGCCGATCATGGGTTGCCCCCAACGAGAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-09-125-619-3
; Sequence 3, Application US/09125619
; Patent No. 6437116
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG

APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: US/99-01-27
; CURRENT APPLICATION NUMBER: US/09/125,619
; CURRENT FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match 7.3%; Score 33.4; DB 4; Length 7766;
Best Local Similarity 47.4%; Pred. No. 0.56;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCGCTTCTCTGCAAGCATGCCCTAAATCTGAAGACCATTAACCCAG 216
DB 4432 GTTACGGCTGCGCTGCTGCTGAGCAGATGAGAGAGAGCTGAGAGCTTAACAAATCCG 4491
QY 217 AGCGTTATGATTAACGCGCGCATGTAGCGCAGGTCGGATATAGTACTATTGAATCG 276
DB 4492 ATTGCTCTCTATTTGGAAGGATTAAGAGATGTCCGATTTTGTGAAGATGAGATG 4551
QY 277 ACTCAGATGTTTCAAGAAATATGCAACATGACCAATGAGCAAGTGAAGCTTAACCTCCGAT 336
DB 4552 AAGAAAGATATCATGATGCTGCTGCTATTGCTTTGAGGGGATGCTTAAGATGAAG 4611
QY 337 ATTACTGTCGGCAATAGCGGATTAATACG 367
DB 4612 TTGCTGTGAAGATTAATGATGTGAGAAAG 4642

RESULT 6
US-09-220-132-30
; Sequence 30, Application US/09220132
; Patent No. 6506607
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE IDENTIFICATION AND ASSESSMENT
; FILE REFERENCE: 07334-074001
; CURRENT APPLICATION NUMBER: US/09/220,132
; CURRENT FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: US 60/079,303
; PRIOR FILING DATE: 1998-03-25
; PRIOR APPLICATION NUMBER: US 60/068,821
; PRIOR FILING DATE: 1997-12-24
; NUMBER OF SEQ ID NOS: 191
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 30
; LENGTH: 14770
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-220-132-30

Query Match 7.3%; Score 33.4; DB 4; Length 14770;
Best Local Similarity 48.7%; Pred. No. 0.8;
Matches 91; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 220 GGTATGTTAAGCGCGGATGTAAGCCAGGTCGGATTAATGACTATTGAATGACT 279
DB 10282 GGAATCTGGAAGAAATATGATTAATTTGGCTTGCTTACCTCAGAAATGAAATGCC 10341
QY 280 CAGATGTTTCAAGAAATATGCAACATGACAGTGAAGCTTAACAACTCGATATT 339

```

/ Patent No. 6610836
/ Application US/09489039A
/
/ GENERAL INFORMATION:
/ APPLICANT: Gary Breton et. al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
/ TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 2709.2004001
/ CURRENT APPLICATION NUMBER: US/09/489,039A
/ PRIOR FILING DATE: 2000-01-27
/ PRIOR APPLICATION NUMBER: US 60/117,747
/ NUMBER OF SEQ ID NOS: 1995-01-29
/ SEQ ID NO 3211
/ LENGTH: 1122
/ TYPE: DNA
/ ORGANISM: Klebsiella pneumoniae
/ US-09-489-039A-3211

```

	357	360	363	366	369	372	375	378	381	384	387	390	393	396	399	402	405	408	411	414	417	420	423	426	429	432	435	438	441	444	447	450	453	456	459	462	465	468	471	474	477	480	483	486	489	492	495	498	501	504	507	510	513	516	519	522	525	528	531	534	537	540	543	546	549	552	555	558	561	564	567	570	573	576	579	582	585	588	591	594	597	600	603	606	609	612	615	618	621	624	627	630	633	636	639	642	645	648	651	654	657	660	663	666	669	672	675	678	681	684	687	690	693	696	699	702	705	708	711	714	717	720	723	726	729	732	735	738	741	744	747	750	753	756	759	762	765	768	771	774	777	780	783	786	789	792	795	798	801	804	807	810	813	816	819	822	825	828	831	834	837	840	843	846	849	852	855	858	861	864	867	870	873	876	879	882	885	888	891	894	897	900	903	906	909	912	915	918	921	924	927	930	933	936	939	942	945	948	951	954	957	960	963	966	969	972	975	978	981	984	987	990	993	996	999	1002	1005	1008	1011	1014	1017	1020	1023	1026	1029	1032	1035	1038	1041	1044	1047	1050	1053	1056	1059	1062	1065	1068	1071	1074	1077	1080	1083	1086	1089	1092	1095	1098	1101	1104	1107	1110	1113	1116	1119	1122	1125	1128	1131	1134	1137	1140	1143	1146	1149	1152	1155	1158	1161	1164	1167	1170	1173	1176	1179	1182	1185	1188	1191	1194	1197	1200	1203	1206	1209	1212	1215	1218	1221	1224	1227	1230	1233	1236	1239	1242	1245	1248	1251	1254	1257	1260	1263	1266	1269	1272	1275	1278	1281	1284	1287	1290	1293	1296	1299	1302	1305	1308	1311	1314	1317	1320	1323	1326	1329	1332	1335	1338	1341	1344	1347	1350	1353	1356	1359	1362	1365	1368	1371	1374	1377	1380	1383	1386	1389	1392	1395	1398	1401	1404	1407	1410	1413	1416	1419	1422	1425	1428	1431	1434	1437	1440	1443	1446	1449	1452	1455	1458	1461	1464	1467	1470	1473	1476	1479	1482	1485	1488	1491	1494	1497	1500	1503	1506	1509	1512	1515	1518	1521	1524	1527	1530	1533	1536	1539	1542	1545	1548	1551	1554	1557	1560	1563	1566	1569	1572	1575	1578	1581	1584	1587	1590	1593	1596	1599	1602	1605	1608	1611	1614	1617	1620	1623	1626	1629	1632	1635	1638	1641	16
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	----

RESULT 8
 5486473-3/c
 ; Patent No. 5486473
 ; APPLICANT: FUKUTA, HIROYUKI, YOSHIDA, IMAO, TAKAGI, MITSUO,
 ; MANABE, SADO, FUKAI, KONOSUE
 ; TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGENE
 ; NUMBER OF SEQUENCES: 16
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/194,049
 ; FILING DATE: 03-FEB-1994
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 809,255
 ; FILING DATE: 18-DEC-1991
 ; APPLICATION NUMBER: 279,685
 ; FILING DATE: 05-DEC-1988
 ; APPLICATION NUMBER: 932,419
 ; FILING DATE: 19-NOV-1986
 ; SEQ ID NO:3
 ; 5486473-3 LENGTH: 1491

QY	178	CAAAGCGATGCCGTTAATCTGAACGACCTTACCGAGACGGTTATGTAAAGCGGCG	237
Db	204	CAAGTCACAGACCGAAGCTAGATTAAACAGTAACTCGACACATCTGCAGAGTTGGCTCTTC	145

```

RESULT 9
US-09-540-236-492/C
; Sequence 492, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATARRHALIS
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 3840
; SEQ ID NO 492
; LENGTH: 2436
; TYPE: DNA
; ORGANISM: M.catarrhalis
US-09-540-236-492

```

Oy	209	TTTACCAGAGCGGTTATGGTAAACGGCCGCAATGTAAGCCAGGGTGGCAATTAATAGTACTA	268
Db	1511	TTTACCACATTTATCATTAATCTTAAAGCAGACTTCGGGCATGGTGGTGAATTAATTAAGC	1455
Oy	269	TTGA---ATGCTCAGAAATGGTTTCGAATAATATGCCACATCGACCGATGGAAAGCGTAT	325
Db	1451	TGAACTTCATATCCAAATATGCTTCAATTAATTAACCGCATGTTCGGCCCTCTCCACAAAGA	1392
Oy	326	AAAACCTCCGATTAATCTGTGGCCAAATACGCGCGTATTAATAGGCC	369
Db	1391	ACTGTCACTTCATGACCATCAACCAAAACTTGGGTTCGATTGCC	1348

```

RESULT 10
US-09-596-002-31/C
; Sequence 31, Application US/09596002
; Patent No. 6632636
GENERAL INFORMATION:
; APPLICANT: Lagace, Robert, E.
; APPLICANT: Patterson, Chandra
; APPLICANT: Berg, Kim, L.
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF MORAXELLA CATARRHALIS GENOME
; FILE REFERENCE: PM-0008-4 US
; CURRENT APPLICATION NUMBER: US/09/596,002
; PRIOR FILING DATE: 2000-06-16
; PRIOR APPLICATION NUMBER: 60/140,121
; NUMBER OF SEQ ID NOS: 1999-06-18
; SOFTWARE:
;         PERL Program
;         41.
; SEQ ID NO 31
; LENGTH: 65792
; TYPE: DNA
; ORGANISM: M. catarrhalis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: incyte template ID No. 6632636 31
; PUBLICATION INFORMATION:
US-09-596-002-31

```

Query Match	6.8%	Score 31.2;	DB 4;	Length 65792
Best Local Similarity	53.7%;	Pred. No. 11;		

Matches 88; Conservative 0; Mismatches 73; Indels 3; Gaps 1;

QY 209 TTACCCAGAGGGTATGTGTACGCGCGCATGTAGCGAGGTCCGATTAATAGTACTA 268
DB 6819 TTACCCAGATTCATCATATCTTAAAGGCAAGTTCGGCATGTCTTATGAATTAAGC 6760

QY 269 TTGA---ACTGACTCAGATGGTTTCAGAAATTAATGACCACCATGACCGAGTGAAGCTA 325
DB 6759 TGAATTCATTAATCCAAATGCTTCATTAATTAAGCCAGATTCGCCCTTGACAGAA 6700

QY 326 AAAACTCCGATATTAATCTGTGCGCCATACGCGGTAAATACGCC 369
DB 6659 ACTGTACCTCATGACCATCAACCAAACTTCGTTGATTCGCC 6656

RESULT 11
US-08-508-761B-1
Sequence 1, Application US/08508761B
Patent No. 6027920
GENERAL INFORMATION:
APPLICANT: Joliff, Gwennael
APPLICANT: Guyonvarch, Arnel
APPLICANT: Purification, Relano
APPLICANT: Duchiron, Francis
APPLICANT: Renaud, Michel
TITLE OF INVENTION: System for Protein Expression and
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Jacobson, Price, Holman & Stern, PLLC
STREET: 400 Seventh St. N.W.
CITY: Washington D.C.
COUNTRY: U.S.A.
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/508,761B
FILING DATE: 31-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09652
FILING DATE: 29-JUL-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 91/09870
FILING DATE: 02-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Player, William E.
REGISTRATION NUMBER: 31,409
REFERENCE/DOCKET NUMBER: P58525NA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 638-6666
TELEFAX: (202) 393-5350
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2547 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: *Corynebacterium Melasecola*
FEATURE:
NAME/KEY: CDS
LOCATION: 420..2390
US-08-508-761B-1

Query Match 6.8%; Score 30.8; DB 3; Length 2547;
Best Local Similarity 58.9%; Pred. No. 2.4;

Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 280 CAGATGTTTCAGAAATATATCCACATGACCAAGTGAAGCTTAAACTCCGATATT 339
DB 876 CAGATGTTTCAGAAATATATCCACATGACCAAGTGAAGCTTAAACTCCGATATT 935

QY 340 ACTGTGCGCCAAATACGCGCGGTAAATACGCC 369
DB 936 GTTGCTCTCCATATGCTGTGCGAGAGCTTC 965

RESULT 12
US-09-103-840A-2/c
Sequence 2, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 2
LENGTH: 4403765
TYPE: DNA
ORGANISM: *Mycobacterium tuberculosis*
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 6.7%; Score 30.6; DB 3; Length 4403765;
Best Local Similarity 51.9%; Pred. No. 95;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 45 TGGCAGTCTCTGCGCGCGGTTCACATGCGCGCGCGGTATCATACGCGCG 104
DB 1357218 TTGCGAGCGCTGCTGCTACGCGCGCGGTCTACACCGCGCGCGGTCTGCA 1357159

QY 105 CGGCAATAGTTCGCGCGCGCACTCAAGCTTGAGCATTTATCAGTACGCGCG 164
DB 1357158 AAACACCGCATGCGGTACGAGCCACGATTAAGGTGTGCGCATGCTGATCTGCGCG 1357099

QY 165 TGGCCTTGCTCTG 177
DB 1357098 CCCGAGCTGCTG 1357086

RESULT 13
US-09-103-840A-1/c
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION:
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
CURRENT FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA

```

; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Query Match
Best Local Similarity 6.7%; Score 30.6; DB 3; Length 4411529;
Matches 69; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 45 TGGCAGTGTCTGCGTGGCGTGTCTTCCACAATGGGGCGCGCGGTATCATTAACGGCGG 104
DB 1357750 TTGCGAGGCGCTGGCTGACGGCGCGCTGCACACGGCGCGCTGTCAGATCCGTGCA 1357691

QY 105 CGGCAATAGTTCGGCGCGGACTCAACGTTGATTCAGTACGCTTCGGCTAACGC 164
DB 1357690 AAACACCGCATGCCCTTAAGAGGCCACGATAGGTGTGCGCATGTGATCTGTGCGGC 1357631

QY 165 TGGCTTGTCTGT 177
DB 1357630 CCGGACGTCGTG 1357618

RESULT 14
US-09-350-756-2/C
; Sequence 2, Application US/09350756
; Patent No. 6495143
; GENERAL INFORMATION:
; APPLICANT: U.S. Army Medical Research Institute for Infectious Diseases
; APPLICANT: John S. Lee
; APPLICANT: Peter Pushko
; APPLICANT: Michael D. Parker
; APPLICANT: Jonathan F. Smith
; APPLICANT: Mark T. Dertzbaugh
; APPLICANT: Leonard Smith
; TITLE OF INVENTION: Botulinum Neurotoxin Vaccine
; FILE REFERENCE: 003/124/SAP R1D 98-21
; CURRENT APPLICATION NUMBER: US/09/350,756
; CURRENT FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: US 60/092,416
; EARLIER FILING DATE: 1998-07-10
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Apple Macintosh Microsoft Word 6.0
; SEQ ID NO 2
; LENGTH: 1341
; TYPE: DNA
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: /note="synthetic sequence encoding botulinum neurotoxin"
US-09-350-756-2
Query Match
Best Local Similarity 6.6%; Score 30; DB 4; Length 1341;
Matches 63; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 330 CTCGATATTACTGTGGCCCAATACGGCGGTATTAAGCCGCGTGTATATATATCA 389
DB 673 CTTGATATTGAGCTGAGACGTTGGGTGTTGAAGATGAGAGATATTTTCATCAGATGA 614

QY 390 GCTGTTACCCGTGTGTTTACCATGAAATGACATGCAACAACGCGCGCTAAC 447
DB 613 ACTGGGTACGATGATGTCAACCGTCCAGTTGAAGATGATTTCAACGTTAGCATTAAC 556

RESULT 15
US-09-252-991A-10904
; Sequence 10904, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
```

```

; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 10904
; LENGTH: 1008
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-10904
Query Match
Best Local Similarity 6.5%; Score 29.8; DB 4; Length 1008;
Matches 61; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 41 TTTCGCAATGCTCTGCGTGGCGTGTCTTCCACAATGGGGCGCGCGGTATCATTAACG 100
DB 141 TGTATGCTGTGTGCTGACCGAGGCGGTGCAAGCGCTTGGCGCGCGAGCCGCTGG 200

QY 101 GCGGCGCAATAGTTCGGCGCGGACTCAACGTTGAGCATTTATCATTAACGAT 153
DB 201 GCGGGAGAGAGAGCCCGACCGGTGCGCGCGTGTGATTAAGCACGAT 253
```

Search completed: March 16, 2004, 04:37:19
Job time : 63.6647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456
Sequence: 1 atgaactttaaagtcgc.....ccacgctaccagctataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 243257 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	34.2	7.5	616	12 US-10-424-599-88945	Sequence 88945, A
C 2	34	7.5	1344	9 US-09-835-684-6	Sequence 6, Appli
C 3	34	7.5	1344	9 US-09-880-371-6	Sequence 6, Appli
C 4	34	7.5	1344	9 US-09-879-248-5	Sequence 5, Appli
C 5	34	7.5	1344	14 US-10-010-390-6	Sequence 5, Appli
C 6	34	7.5	1344	15 US-10-441-736-5	Sequence 6, Appli
C 7	34	7.5	2886	9 US-09-801-368-131	Sequence 11, App
C 8	33.6	7.4	2751	15 US-10-398-221-3563	Sequence 3563, Ap
C 9	33.6	7.4	3532	14 US-10-017-161-1913	Sequence 1913, Ap
C 10	33.4	7.3	3532	15 US-10-292-798-1569	Sequence 1569, Ap
C 11	33.4	7.3	7766	12 US-10-222-566-3	Sequence 3, Appli
C 12	33.4	7.3	7766	14 US-10-222-162-3	Sequence 3, Appli
C 13	33.4	7.3	7766	14 US-10-143-024-3	Sequence 3, Appli
C 14	33.4	7.3	14770	10 US-09-873-367C-230	Sequence 230, App
C 15	33.2	7.3	1355	15 US-10-045-674-593	Sequence 593, App

C 16	33.2	7.3	1662	14 US-10-156-761-3483	Sequence 3483, Ap
C 17	33.2	7.3	9025608	14 US-10-156-761-1	Sequence 1, Appli
C 18	32.2	7.1	568	14 US-10-333-631-1	Sequence 1, Appli
C 19	32.2	7.1	972	14 US-10-333-631-1	Sequence 4, Appli
C 20	32.2	7.1	1110	12 US-10-282-122A-23637	Sequence 23637, A
C 21	32.2	7.1	4830	12 US-10-282-122A-27869	Sequence 27869, A
C 22	32	7.0	1776	12 US-10-282-122A-24385	Sequence 24385, A
C 23	31.8	7.0	978	15 US-10-369-493-24265	Sequence 24265, A
C 24	31.8	7.0	24081	14 US-10-132-134-13	Sequence 13, Appli
C 25	31.8	7.0	52101	14 US-10-132-134-1	Sequence 1, Appli
C 26	31.6	6.9	6003	13 US-10-010-901-9	Sequence 9, Appli
C 27	31.4	6.9	840	12 US-10-282-122A-31900	Sequence 31900, A
C 28	31.4	6.9	1245	15 US-10-369-493-40803	Sequence 40803, A
C 29	31.2	6.8	2394	12 US-10-282-122A-26837	Sequence 26837, A
C 30	31	6.8	525	10 US-09-918-995-32144	Sequence 32144, A
C 31	30.8	6.8	715	12 US-10-424-599-10132	Sequence 10132, A
C 32	30.8	6.8	725	9 US-09-919-580-878	Sequence 878, App
C 33	30.8	6.8	2050	14 US-10-086-510-1	Sequence 1, Appli
C 34	30.8	6.8	2699	15 US-10-260-238-418	Sequence 418, App
C 35	30.6	6.7	1164	12 US-10-282-122A-13500	Sequence 13500, A
C 36	30.6	6.7	2301	9 US-09-815-242-844	Sequence 4844, Ap
C 37	30.6	6.7	2301	9 US-09-815-242-9042	Sequence 9042, App
C 38	30.6	6.7	75216	14 US-10-080-170-646	Sequence 646, App
C 39	30.4	6.7	392	12 US-10-282-122A-25231	Sequence 25231, A
C 40	30.4	6.7	992	12 US-10-425-114-31747	Sequence 31747, A
C 41	30.4	6.7	1024	15 US-10-323-069A-111	Sequence 111, App
C 42	30.4	6.7	1601042	15 US-10-027-632-55064	Sequence 55064, A
C 43	30.2	6.6	654	12 US-10-425-114-28546	Sequence 28546, A
C 44	30.2	6.6	829	15 US-10-027-632-164999	Sequence 164999, A
C 45	30.2	6.6	1203	9 US-09-738-626-1184	Sequence 3184, Ap

ALIGNMENTS

RESULT 1
US-10-424-599-88945/C
Sequence 88945, Appli
Publication No. US20040031072A1
GENERAL INFORMATION:
APPLICANT: La Rosa Thomas J
APPLICANT: Kovalic David K
APPLICANT: Zhou Yihua
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(53223)B
CURRENT APPLICATION NUMBER: US/10/424,599
CURRENT FILING DATE: 2003-04-28
NUMBER OF SEQ. ID NOS: 285684
SEQ ID NO 88945
LENGTH: 616
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(616)
OTHER INFORMATION: unsure at all n locations
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT1847_51326C.1
US-10-424-599-88945

Query Match 7.5%; Score 34.2; DB 12; Length 616;
Best Local Similarity 50.9%; Pred. No. 0.51;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 101 GCGCGGCAATGATGTCGCGCGGACATCAAGTTAGCATTTATCAGTACGTTCCGCTA 160
Db 556 GGGCCCAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497
QY 161 ACAGTGGCTGCTCTGCAAGGATGATGATGATGATGATGATGATGATGATGATGATG 220
Db 496 ATAGTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 437

QY 221 GTTATGTAACGCCCGCATGTAGCCAGGATGAGGATA 259
 Db 436 GTAATGTAAGGATGTAATGCAAGGAGGATTCGGGTA 398

RESULT 2
 US-09-835-684-6/c
 ; Sequence 6, Application US/09835684
 ; Patent No. US20020019337A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Oiu, Dewen
 ; APPLICANT: Remick, Dean
 ; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
 ; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
 ; FILE REFERENCE: 21829/71
 ; CURRENT APPLICATION NUMBER: US/09/835,684
 ; PRIOR FILING DATE: 2001-04-16
 ; PRIOR APPLICATION NUMBER: 60/198,359
 ; NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-835-684-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGGCTGGCGCTGCTTCACAAATGGGCGCGCGCGTATCATTAACGCGCGGCAAT 111
 Db 209 GCTCCGGTGGCGCGCATTCATGATTTGGCGATAGACGACTTTAAAGTTCCGCCAAT 150
 QY 112 AGTTCGCGCGCGGACTCAACGTTGAGCATTATCACTAGCAGTTCCGCTAAGCGCGCTT 171
 Db 149 AATTGAGCCATTGCTCAATGTTGCCATGATGAGGTTGTGCCCAACGCAAAATTT 90
 QY 172 GC 173
 Db 89 GC 88

RESULT 3
 US-09-880-371-6/c
 ; Sequence 6, Application US/09880371
 ; Patent No. US20020059658A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Derocher, Jay
 ; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
 ; FILE REFERENCE: 21829/91
 ; CURRENT APPLICATION NUMBER: US/09/880,371
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: 60/211,585
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-880-371-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGGCTGGCGCTGCTTCACAAATGGGCGCGCGGTAATCATTAACGCGCGGCAAT 111
 Db 209 GCTCCGGTGGCGCGCATTCATGATTTGGCGATAGACGACTTTAAAGTTCCGCCAAT 150
 QY 112 AGTTCGCGCGCGGACTCAACGTTGAGCATTATCACTAGCAGTTCCGCTAAGCGCGCTT 171
 Db 149 AATTGAGCCATTGCTCAATGTTGCCATGATGAGGTTGTGCCCAACGCAAAATTT 90
 QY 172 GC 173
 Db 89 GC 88

RESULT 4
 US-09-879-248-5/c
 ; Sequence 5, Application US/09879248
 ; Patent No. US20020062500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fan, Hao
 ; APPLICANT: Wei, Zhong-Min
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
 ; FILE REFERENCE: 21829/81
 ; CURRENT APPLICATION NUMBER: US/09/879,248
 ; PRIOR FILING DATE: 2001-06-12
 ; PRIOR APPLICATION NUMBER: 60/212,211
 ; NUMBER OF SEQ ID NOS: 18
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 5
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-879-248-5

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGGCTGGCGCTGCTTCACAAATGGGCGCGCGGTAATCATTAACGCGCGGCAAT 111
 Db 209 GCTCCGGTGGCGCGCATTCATGATTTGGCGATAGACGACTTTAAAGTTCCGCCAAT 150
 QY 112 AGTTCGCGCGCGGACTCAACGTTGAGCATTATCACTAGCAGTTCCGCTAAGCGCGCTT 171
 Db 149 AATTGAGCCATTGCTCAATGTTGCCATGATGAGGTTGTGCCCAACGCAAAATTT 90
 QY 172 GC 173
 Db 89 GC 88

RESULT 5
 US-10-010-390-6/c
 ; Sequence 6, Application US/10010390
 ; Publication No. US20030104979A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Oviedo, Ernesto
 ; TITLE OF INVENTION: METHODS OF INHIBITING DESICCATION OF CUTTINGS REMOVED
 ; FILE REFERENCE: 21829/111
 ; CURRENT APPLICATION NUMBER: US/10/010,390
 ; PRIOR FILING DATE: 2001-11-05
 ; PRIOR APPLICATION NUMBER: 60/248,169
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora

Dy 238 GATGTAGCCAGGGTCCGGAATAAGTACTATTGAACTGACAGATGGTTTCAGAAT 297
| | | | |
Db 321 GAGTAGTGAAAGAGCAAAAATTACTATCGAACCGTTTCTCGTAGCATTAT 380

Query Match 7.3%; Score 33.4; DB 12; Length 7766;

Best Local Similarity 47.4%; Pred. No. 4;

Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATGCCGTAATCTGAAGACCACTTACCAG 216
4432 GTTACGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCGACAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGATAGCCAGGCTGCGATATAGTACTATTGAACGTG 276
4492 ATTCTGCTGCTATTTGGAGAGGGTATGAGATGGTGGGATTTTGTAGAGATGAGATG 4551
277 ACTAGATGTTTCAAGAAATATATGACCAATGACCACTGGAACGCTTAAACTCCGAT 336
4552 AAGAAAGATGATCAAGATTGCTGCTGCTATTTGAGAGGGATGCTTAAGAGATGAAAG 4611
337 ATTACTGTGGCCCAATACGCGGCTAATAACG 367
4612 TTGCTGTGAAGAGATGATGATGTGAGAAAG 4642

RESULT 12

US-10-222-162-3

Sequence 3, Application US/10222162

Publication No. US20030060618A1

GENERAL INFORMATION:

APPLICANT: NORRIS, STEVEN J.

APPLICANT: JING-REN, ZHANG

APPLICANT: HARDHAM, JOHN M.

APPLICANT: HOWELL, JERRILYN K.

APPLICANT: BARBOUR, ALAN G.

APPLICANT: WEINSTOCK, GEORGE M.

TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA

FILE REFERENCE: UTS# 234USD4

CURRENT FILING DATE: 2002-08-16

PRIOR APPLICATION NUMBER: 09/125,619

NUMBER OF SEQ ID NOS: 50

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 7766

TYPE: DNA

ORGANISM: Borrelia burgdorferi

FEATURE:

NAME/KEY: modified_base

LOCATION: (127)

OTHER INFORMATION: R = A OR G

US-10-222-162-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;

Best Local Similarity 47.4%; Pred. No. 4;

Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATGCCGTAATCTGAAGACCACTTACCAG 216
4432 GTTACGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCGACAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGATAGCCAGGCTGCGATATAGTACTATTGAACGTG 276
4492 ATTCTGCTGCTATTTGGAGAGGGTATGAGATGGTGGGATTTTGTAGAGATGAGATG 4551
277 ACTAGATGTTTCAAGAAATATATGACCAATGACCACTGGAACGCTTAAACTCCGAT 336
4552 AAGAAAGATGATCAAGATTGCTGCTGCTATTTGAGAGGGATGCTTAAGAGATGAAAG 4611
337 ATTACTGTGGCCCAATACGCGGCTAATAACG 367
4612 TTGCTGTGAAGAGATGATGATGTGAGAAAG 4642

RESULT 13

US-10-143-024-3

Sequence 3, Application US/10143024

Publication No. US20030092903A1

GENERAL INFORMATION:

APPLICANT: NORRIS, STEVEN J.

APPLICANT: JING-REN, ZHANG

APPLICANT: HARDHAM, JOHN M.

APPLICANT: HOWELL, JERRILYN K.

APPLICANT: BARBOUR, ALAN G.

APPLICANT: WEINSTOCK, GEORGE M.

TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA

FILE REFERENCE: UTS# 234USD1

CURRENT FILING DATE: 2002-05-10

PRIOR APPLICATION NUMBER: 09/125,619

NUMBER OF SEQ ID NOS: 48

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 3

LENGTH: 7766

TYPE: DNA

ORGANISM: Borrelia burgdorferi

FEATURE:

NAME/KEY: modified_base

LOCATION: (127)

OTHER INFORMATION: R = A OR G

US-10-143-024-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;

Best Local Similarity 47.4%; Pred. No. 4;

Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

157 GCTAACGCTGGCTGCTCTGCAAGCGATGCCGTAATCTGAAGACCACTTACCAG 216
4432 GTTACGCTGGCGCTGCTGAGACAGATGAGAGAAAGCTGCGACAGGCTACCAATCCG 4491
217 AGCGTTATGTTAAGCGCGCGATGATAGCCAGGCTGCGATATAGTACTATTGAACGTG 276
4492 ATTCTGCTGCTATTTGGAGAGGGTATGAGATGGTGGGATTTTGTAGAGATGAGATG 4551
277 ACTAGATGTTTCAAGAAATATATGACCAATGACCACTGGAACGCTTAAACTCCGAT 336
4552 AAGAAAGATGATCAAGATTGCTGCTGCTATTTGAGAGGGATGCTTAAGAGATGAAAG 4611
337 ATTACTGTGGCCCAATACGCGGCTAATAACG 367
4612 TTGCTGTGAAGAGATGATGATGTGAGAAAG 4642

RESULT 14

US-09-873-367C-230

Sequence 230, Application US/09873367C

Publication No. US20030165839A1

GENERAL INFORMATION:

APPLICANT: Young, Paul

APPLICANT: Soppec, Daniel

APPLICANT: Endress, Gregory

APPLICANT: Augustus, Meena

APPLICANT: Ebner, Reinhard

APPLICANT: Carter, Kenneth

TITLE OF INVENTION: Signature Gene Sets

FILE REFERENCE: 689290-64

CURRENT FILING DATE: 2003-04-29

PRIOR APPLICATION NUMBER: US/09/873,367C

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: U.S. 60/236,891

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: U.S. 60/236,842

PRIOR FILING DATE: 2000-09-29

PRIOR APPLICATION NUMBER: U.S. 60/244,867

PRIOR FILING DATE: 2000-11-01

PRIOR APPLICATION NUMBER: U.S. 60/245,084

PRIOR FILING DATE: 2000-11-01

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456
Sequence: 1 atgaacttttaaaagtggc.....ccacgcgtacacagatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estba:*
2: em_esthum:*
3: em_estcin:*
4: em_estcnu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estcro:*
8: em_estc1:*
9: gb_estc1:*
10: gb_estc2:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estctun:*
16: em_estcom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vrc:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_phg:*
27: em_ges_vr1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216.6	47.5	680	12	BU618688 BU618688
2	35.2	7.7	813	28	B2204853 BU230-390
3	34.8	7.6	801	12	B1648266 BU3278209
4	34.2	7.5	523	12	BU334624 BU334624

5	34.2	7.5	1036	12	BG671547 DBRBUB05
6	34	7.5	234	10	BB581987
7	34	7.5	500	14	CA711477
8	34	7.5	811	28	B2469000
9	33.8	7.4	296	9	AV058630
10	33.8	7.4	365	9	AA746477
11	33.8	7.4	558	14	CD374421
12	33.8	7.4	559	13	BU765156
13	33.8	7.4	695	28	AQ637195
14	33.8	7.4	941	29	CG178280
15	33.8	7.4	1098	29	CNS02GQP
16	33.6	7.4	374	12	BY236219
17	33.6	7.4	594	13	B1335653
18	33.6	7.4	653	12	B1336964
19	33.6	7.4	712	14	CF437982
20	33.4	7.3	505	6	AL809250
21	33.4	7.3	624	14	CB578319
22	33.4	7.3	664	13	BX506928
23	33.4	7.3	679	29	CC961526
24	33.4	7.3	695	28	BH633951
25	33.4	7.3	699	12	B1959328
26	33.4	7.3	700	28	BH651371
27	33.4	7.3	709	28	BH690914
28	33.4	7.3	715	13	BU444683
29	33.4	7.3	785	28	BH578793
30	33.4	7.3	815	28	BH461716
31	33.4	7.3	827	28	BH438448
32	33.4	7.3	830	28	BH715216
33	33.4	7.3	835	28	BH575194
34	33.4	7.3	853	28	BH578402
35	33.4	7.3	856	14	CD456845
36	33.4	7.3	1026	28	BZ458784
37	33.2	7.3	589	28	BZ300068
38	33.2	7.3	592	29	CG072975
39	33.2	7.3	629	9	AU295325
40	33.2	7.3	675	28	AQ361497
41	33.2	7.3	788	9	AU139321
42	33.2	7.3	1646	10	BF579120
43	33	7.2	456	12	BU331339
44	33	7.2	524	12	BU367372
45	33	7.2	547	12	BU333958

ALIGNMENTS

RESULT 1
BU618688/c
LOCUS
DEFINITION
BU618688 NIBB Mochii normalised Xenopus early gastrula library
Xenopus laevis cDNA clone XL186D22 5', mRNA sequence.
BU618688
BU618688.1 GI:37256713

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

REFERENCE

Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and
Kohata, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp
The information of this clone is available through the following
URL.

http://xenopus.nibd.ac.jp

FEATURES

Location/Qualifiers

1. 680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL186b2"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_1db="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match

Best Local Similarity 47.5%; Score 216.6; DB 12; Length 680;
Matches 306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTAAAGTGGAGCGCTTGGCAAGTGTGTTCTGGGAGTGGCTGGCT
DB 593 ATGAACTTTAAAGTGGAGCGCTTGGCAAGTGTGTTCTGGGAGTGGCTGGCT
QY 61 GGCGTCTTCCACATGGGGCGGGCGGTAATCAATACGGCGCGCAATAGTCCGGC
DB 533 GGTGTGTCTCTCAGTACGGCGCGGCTAACCGCGTGTGGCGTAATAATAGCGGC
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCTTGGCTGGCA
DB 473 CCAAAATTCTGAGTGAACTTTTACAGTACGTTGGCGGTAACTTGCCTCTGCAA
QY 181 AGGATGCGCGGTAATCTGAAACGACATTAACCGAGCGGTTATGTAAAGCGCGCAT
DB 413 ACTGATGCCCGTACTGACTGACTGACTGACTGACTGACTGACTGACTGACTGACTG
QY 241 GTAGGCGAGGCGCGGATATGATGATGATGATGATGATGATGATGATGATGATGAT
DB 353 GTTGTGTAGGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 301 GCCACCATGACGCGTGGAGAGCTTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA
DB 293 GCTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 361 AATAAGCCCGCGCTGTTATATATGATGATGATGATGATGATGATGATGATGATGAT
DB 233 GCGAACGCGTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
QY 421 GCACATGCAACACGCGCGCTAACCGATGATGATGATGATGATGATGATGATGATGAT
DB 173 GCCTTGTGTAACGCGCGCTCATCATCATCATCATCATCATCATCATCATCATCAT

```

RESULT 2

BZ204853

LOCUS

DEFINITION

CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone

VERSION

BZ204853.1 GI:23862905

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 813)

Zhaio, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K.,

Shivatsbeyn, A., Gebregiorgis, E., Overton, L., Russell, D., Chen, D.,

Rat BAC End Sequences from Library CHORI-230 Mbol segment

Other-SSS: CH230-390D21.TV

Contact: Shaying Zhao

The Institute for Genomic Research

FEATURES

Location/Qualifiers

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cno.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or erting_information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Seq primer: 77
Class: BAC ends.

ORIGIN

Query Match

Best Local Similarity 7.7%; Score 35.2; DB 28; Length 813;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

```

QY 178 CAAGCGATGCCGTAATGTAAGACATTAACCGAGCGGTATGTAACGGCGCC
DB 548 CAGAGCTTCACATCAATCAACAGACACTCAACTTAATGAAGAAACATGAGGAACA
QY 238 GATTAAGCCGAGGTCGATATGATGATGATGATGATGATGATGATGATGATGAT
DB 608 TCTGGAACCATGAGGACGAGGAAAAATCTCTTAACCAACCAATGGCTTATGCTC
QY 298 AATGCCACATGACGACGAGGAAAGCTTAAAGTAAAGTAAAGTAAAGTAAAGTAA
DB 668 AAGATCAATGACCAATGAGATCTCATTAACGCAAACTTCTGTAGCAAAAGGACAC
QY 358 GGTATTAACGCC 369
DB 728 TGTGTTAGGAC 739

```

RESULT 3

B1648266

LOCUS

DEFINITION

603278209P1 NCT CGAP Mam3 Mus musculus cDNA clone IMAGE:5318307 5'

VERSION

B1648266.1 GI:15562502

KEYWORDS

SOURCE

ORGANISM

Mus musculus (house mouse)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 801)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cga@bbs-remail.nih.gov

Tissue Procurement: lotar Hennighausen Ph.D., Chu-Xia Deng Ph.D.

cDNA Library Preparation: Life Technologies, Inc.

DNA Sequencing by: InCyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/BLND at:

http://image.llnl.gov
Plate: LLM11807 row: b column: 04
High quality sequence stop: 798.
Location/Qualifiers

FEATURES

Source

1. .801
/organism="Mus musculus"
/mol_type="mRNA"
/strain="129, C57BL/6J, FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:5318307"
/tissue_type="tumor, gross tissue"
/dev_stage="10 months"
/lab_host="DH10B"
/note="Organ: mammary, Vector: pCMV-Sport6; Site 1: SalI; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Library constructed by Life Technologies. Investigators providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH Reference for transgenic model: Xu et al., Nature Genetics 22, 37-43 (1999)."

ORIGIN

Query Match 7.6%; Score 34.8; DB 12; Length 801;
Best Local Similarity 50.0%; Pred. No. 76;
Matches 114; Conservative 0; Mismatches 112; Indels 2; Gaps 1;
OY 210 TACCAGAGCGGTTATGTGTAACGGCCGATGTAGCCAGGCGGATATAGTACTAT 269
DB 322 TTCACCCGCAAGCATGTGACCGGTTATCCCTCGGGGCGGCACTGACCGC 381
OY 270 TGAATGCACGATGTTTCAGAA--ATAATGCCACATGCACCACTGGACGCTTAA 327
DB 382 TGAACCTTCACAGATCATGACATGAGAAAGTCCCAAGCCAGAACCCAGC 441
OY 328 AACTCCGATATTACTGTGCGCAATACGCGCGGTATTAATTTATGAT 387
DB 442 GACGCCACCTCGCAATAGCAAGACCGGCTGAGCTCTGCTGCTGCAATGATGAG 501
OY 388 CAGCTGCTTACCCGTTGTTATCCATGAATGGACATGCAACAC 435
DB 502 CTGCTGGTCCGCGATGAGAAAGTTGAGACCCACAGACGAGAGAC 549

RESULT 4

LOCUS

Bj334624 523 bp mRNA linear EST 05-MAR-2002
Bj334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
discoideum cDNA clone dda47006 5', mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (2002)
Contact: Tadabu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yatai, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
Full length cDNA of Dictyostelium discoideum at the aggregation
stage
Location/Qualifiers

FEATURES

Source

1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47006"

/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 93;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
OY 238 GATGAGCCAGGCTGGGATTAATGTAATTTGAAGTACTCAGATGTTTCAGAAAT 297
DB 398 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 457
OY 298 AATGCCACATCGACCGATGAGCGTAAAAATCCGATATTAAT 342
DB 458 AATAACAAACAACAACAATTAATTAATTAATTAATTAATTAAT 502

RESULT 5

LOCUS

BG671547 1036 bp mRNA linear EST 30-APR-2001
DRNBUB05 Rat DRG Library Rattus norvegicus cDNA clone DRNBUB05 5',
mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

COMMENT

1 (bases 1 to 1036)
Xiao, H.S., Huang, Q.H., Zhang, F.X., Bao, L., Lu, Y.J., Guo, C.,
Yang, L., Huang, W.J., Pu, G., Xu, S.H., Cheng, X.P., Yan, Q., Zhu, Z.D.,
Zhang, X., Chen, Z., Han, Z.G. and Zhang, X.
Identification of gene expression profile of dorsal root ganglion
in the rat peripheral axotomy model of neuropathic pain
Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
22056133
JOURNAL
MEDLINE
PUBMED
Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome
Center at Shanghai, 351 Guo Shoujing Road, Zhongjiang Hi-Tech Park,
Pudong New Area, P.R.China. Please contact with Zhang Xu
(xu.zhang@ion.ac.cn) or Han Zenguang (hanzg@chgc.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.
Location/Qualifiers

FEATURES

Source

1. .1036
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DRNBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"

ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.3e+02;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

/clone="wtk2c.pk014.o3"
/tissue_type="kernel"
/clone_lib="wtk2c"
/note="Vector: pBluescript SK+, Site 1: EcoRI, Site 2:
XhoI; wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN

Query Match 7.5%; Score 34; DB 14; Length 500;
Best Local Similarity 54.5%; Pred. No. 1e+02;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 15 AGTGGACGATTCGACGATCTGAGTCTGCTGGCTGGCTGTTCCACA 74
Db 274 AGTGGATGACATCGACGACGCGGCTCCCGCGCTTTCGAACCGCGGAGCTCCCN 215
Qy 75 ATGGGGCGCGCGGATATCATCAACGCGCGGCAATGTTCCGCGCGAC 126
Db 214 TGTGGAGGCGGTGGGATGACATGCGCGCGCGCGCGGTGGCGCCAC 163

RESULT 8
B246900/c 811 bp DNA linear GSS 13-DEC-2002
LOCUS BOOAO58TF BO.1.6.2 KB tot Brassica oleracea genomic clone BOOAO58,
DEFINITION genomic survey sequence.
ACCESSION B246900
VERSION B246900.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 811)
TOWN,C.D., Van Aken,S., Uteback,T., Koo,H. and Frazer,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOOAO58TR
Contact: Chris Town

REFERENCE 9712 Medical Center Drive, Rockville, MD 20850, USA.
Tol: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TF
Class: sheared ends.

FEATURES

SOURCE

Location/Qualifiers
1..811
/organism="Brassica oleracea"
/mol_type="genomic DNA"
/strain="T01000DH3"
/db_xref="taxon:3712"
/clone="BOOAO58"
/clone_lib="BO.1.6.2 KB tot"
/note="Vector: pHOSt; Site 1: BstXI, 1.6-2 kb sheared
total DNA inserted into pHOSt using BstXI linkers"

ORIGIN

Query Match 7.5%; Score 34; DB 28; Length 811;
Best Local Similarity 48.9%; Pred. No. 1.3e+02;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

Qy 176 TCGAAGCGATCCGTAATCTGAAAGACCATTAACCAAGCGTTATGTAACGGC 235
Db 610 TGTAGGCAAGCATGTAATACTGTGTTCCCATTTATCTACATGTTATGACAAATGCT 551
Qy 236 CCGATGAGGCGAGGCGGATATAGTACTATTTGAACGACACGAATGTTTCGAA 295
Db 550 TTGATCTTAATTCATCTGATGTTGACGCTCTCTCTCTCTGAGGATAGTTTAAT 491
Qy 296 ATATGCCACCATGACACGATGAAAGCTAATAATCCGATATTACTGTGGCCAAATACG 355

Db 490 ACTTGTGCACATTCATTAATGAAAAATCGAAATACCTGATTAACCTATTCAACCA 431
Qy 356 GCGGTA 361
Db 430 AAGATA 425

RESULT 9
AV058630 296 bp mRNA linear EST 23-JUN-1999
LOCUS AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA
DEFINITION clone 1810054H02, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 296)
Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Ichikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., Nitsuma,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Suganara,Y., Suzuki,H., Suzuki,H., Tateo,M., Tomaru,Y.,
Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokote,T.,
Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
RIKEN Mouse ESTs
Unpublished (1999)
Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-resortc.riken.go.jp
Thermolabile and thermostable activation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.riken.go.jp) for
further details.

TITLE
JOURNAL
COMMENT

FEATURES
SOURCE
Location/Qualifiers
1..296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"

ORIGIN

Query Match 7.4%; Score 33.8; DB 9; Length 296;
Best Local Similarity 53.4%; Pred. No. 93;
Matches 71; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 251 GTGGGATTAATGATCTATGAACTGACTAGAAATGTTTCAGAAATATGACCACTCG 310
Db 13 GTGGAGACCAAGATTTTCTAGGTCAACAAAGCTGGTAAAGAAAGAAAGCATCACTG 72
Qy 311 ACAGTGAAGCGTAAAGATCCGATTTACTCTCGGCAATACGCGGTAATACGCGC 370
Db 73 TCCAGTGAAGCAAAAGAAAGAAAGAAATATATTAACCAAGATTGAAGTAG 132
Qy 371 CGCTGTTAATTA 383
Db 133 AGATGCCAAATA 145

CD374421/c	LOCUS	558 bp	mRNA	linear	EST 19-LOC
DEFINITION	TNMBmfC2H33	Wuchereria bancrofti	microfilaria	cDNA	(SAM955JL-WBMF)
ACCESSION	Wuchereria bancrofti	CDNA clone	TNMBmfC2H3 5'	mRNA	sequence.
VERSION	CD374421.1	GI:31229939			
KEYWORDS	EST.				
SOURCE	Wuchereria bancrofti				
ORGANISM	Wuchereria bancrofti				
	Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea;				
	Onchocercidae; Wuchereria.				
REFERENCES	1 (bases 1 to 558)				
AUTHORS	Ndi,J.S., Ribeiro,J.M. and Nutman,T.B.				
TITLE	Toward the characterization of the Wuchereria bancrofti microfilarial transcriptome with comparisons to those of Brugia malayi and Onchocerca volvulus				
JOURNAL	Unpublished (2003)				
COMMENT	Contact: Thomas B. Nutman				
	Laboratory of Parasitic Diseases				
	NIH				
	Building 4 Room 126, Bethesda, MD 20892-0425, USA				
	Email: tnutman@niaid.nih.gov				
	Seq primer: PBuescript T3.				
	Location/Qualifiers				
	1..558				
FEATURES					
source	/organism="Wuchereria bancrofti"				
	/mol_type="mRNA"				
	/db_xref="taxon:6293"				
	/clone="TNMBmfC2H3"				
	/dev_stage="microfilaria"				
	/lab_host="XLI-Blue MRP"				
	/clone_idb="Wuchereria bancrofti microfilaria cDNA (SAM955JL-WBMF)"				
	/note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2: Xho I; Lymphatic filarial nematode parasite of humans. mRNA was prepared from approximately 85,000 microfilariae isolated from the blood of an infected individual from Guyana, South America and converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA pol I. The library has 2.9 x 10E5 independent recombinants and the average insert size is ~1kb. The library was constructed by Sandra J. Laney. The library is available from Dr.S.A.Williams, email: genome@smith.edu."				
ORIGIN					
	Query Match	7.4%	Score 33.8;	DB 14;	Length 558;
	Best Local Similarity	60.7%;	Fred. No. 1.2e+02;		
	Matches 65;	Conservative 0;	Mismatches 40;	Indels 2;	Gaps 1;
QY	76 TGGGGCGGCGCGGTATCATATACGCGCGGCAATAGTCCGGCCCGAGCTCAACGTTG	135			
Db	161 TGGCGCGGTGGCGGTACCCGCGTGGTGGCGGTATATATACCGGCCCAATTCTTANGCT	102			
QY	136 AGCATTTATC-AGTACGGTTCGCTACAGCGCTGCGGCTTGCTGCAG	180			
Db	101 GACATTATNCCAGANCGGTGGCGGAGNANCTGCNCTGCTCTGANA	55			
RESULT 12					
LOCUS	BU765156/c	559 bp	mRNA	linear	EST 10-OCT-2002
DEFINITION	saas2ef11.y1 Gm-c1080	Glycine max	cdna clone	SOYBEAN CLONE ID:	
	Gm-c1080-4869 5'	similar to	SV:DI10	ARATH Q00874	
	DNA-DAMAGE-REPAIR/TOLERATION	PROTEIN DRT100	PRECURSOR	;	mRNA
ACCESSION	BU765156				
VERSION	BU765156.1	GI:23733884			
KEYWORDS	EST.				
SOURCE	Glycine max (soybean)				
ORGANISM	Glycine max				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 559)	Shoemaker, R., Keim, P., Vocklin, L., Erpelting, J., Coryell, V., Khanna, A., Bell, B., Marra, M., Hillier, U., Kueba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Pearson, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCan, R., Waterston, R. and Wilson, R.	Public Soybean EST Project	Unpublished (1999)	
	Contact: Shoemaker R/Public Soybean EST Project			
	Public Soybean EST Project			
	Washington University School of Medicine			
	444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA			
	Tel: 314 286 1800			
	Fax: 314 286 1810			
	Email: ees@watson.wustl.edu			
	This clone is available through: Reggen, Invitrogen Corp, 2130			
	South Memorial Parkway Huntsville, AL 35801 For further information			
	call: (800)-533-4363 or contact: cu@reggen.com web site:			
	www.reggen.com			
	Seq primer: -40RP from Gibco			
	High quality sequence stop: 448.			
	Location/Qualifiers			
	1..559			
	/organism="Glycine max"			
	/mol_type="mRNA"			
	/db_xref="taxon:3847"			
	/clone="SOYBEAN CLONE ID: Gm-c1080-4669"			
	/tissue_type="Roots of 8 day old 'Bragg' supernodulating			
	mutant NTS382 seedlings"			
	/dev_stage="8 days old"			
	/lab_host="RD10B"			
	/clone_lib="Gm-c1080"			
	/note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:			
	XhoI; The mRNA was isolated from roots of 8 day old			
	'Bragg' supernodulating mutant NTS382 seedlings that were			
	infected with Bradyrhizobium japonicum, strain USDA 110, 72			
	hours prior to harvest. Dr. Gary Stacey generously donated			
	the tissue. The roots were flash-frozen in liquid			
	nitrogen. Stragene's cDNA Synthesis Kit (catalog number			
	200401) was used to synthesize the cDNA. First-strand			
	synthesis was performed with 5-methyl dCTP, hence the			
	ligated cDNA was hemimethylated. A modification of			
	Stragene's first-strand synthesis primer was used. An			
	'anchor' nucleotide (V=A, C, or G) was added to the 3' end			
	of the primer (GAGAGAGAGAGAGAGAACTGCTCGAG(T)18V) to			
	anchor the primer at the 5' end of the poly(A) tract.			
	After second-strand synthesis, the cDNA ends were filled			
	in with cloned Pfu DNA, ligated to EcoRI adapters and			
	subsequently phosphorylated. The cDNA was then			
	precipitated and redissolved in sterile, RNase-,			
	DNase-free water. The XhoI site within the first-strand			
	synthesis primer was then restricted by digestion with			
	XhoI from Promega (40U/ul); all XhoI sites in the cDNA			
	would be protected by their hemimethylated status. The			
	cDNA constructs were size-fractionated with a 500bp			
	cutoff, using Sephacryl S-500 High Resolution (Pharmacia			
	Biotech) in a 2-cm diameter column and a bed volume of			
	approximately 1ml. The column eluent was precipitated,			
	redissolved, and ligated into Stragene's pBluescript II			
	XR Predigested vector (pBluescript II SK+) vector that			
	has been digested with EcoRI and XhoI, and phosphorylated			
	by Stragene). This library was constructed in the			
	laboratory of Dr. Paul Keim and Dr. Virginia H. Coryell at			
	Northern Arizona University."			

Oy	76	TGGGCGCGCGCGCGGTATTCATTAACGGGGCGGCAATAGTTCGGCCGGACTCAACGTTG	135
Dd	233	TGGGGTGCAGACGCCACCCCATACCAAGCTCCGGCAACAGTACAGTCCGACCATATGTG	174
Oy	136	AGCAATTATATCAGTACGGTTCGGCTACAGCTGCGCTTGCTGTGCAAAAGCATGCCGT	192
Dd	173	AAGATGCCCAAGTAGGGTTCGATCAGGGGACAGCTTTGAAGGCCAAGAGTGTCCCGT	117
RESULT 13			
AO637195/c		695 bp	DNA linear
LOCUS		695 bp	DNA linear
DEFINITION		RPCI-11-481m9.TV RPCI-11 Homo sapiens genomic clone RPCI-11-481m9, genomic survey sequence.	
ACCESSION		AO637195	
VERSION		AO637195.1	GI:5099830
KEYWORDS		GSS.	
ORGANISM		Homo sapiens (human)	
SOURCE		Homo sapiens	
REFERENCE		Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 695)	
AUTHORS		Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and Venter,J.C.	
TITLE		Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building	
JOURNAL		Published (1997)	
COMMENT		Other GSSs: RPCI-11-481m9.TV Contact: Shaying Zhao, William Nierman, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850 Tel: 301 838 0200 Fax: 301 838 0208 Email: hbe@tigr.org Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@jeng.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html. Seq primer: T7 Class: BAC ends.	
FEATURES			
source			
location/Qualifiers			
1..695			
/organism="Homo sapiens"			
/mol_type="genomic DNA"			
/db_xref="GDB:7684616"			
/db_xref="taxon:9606"			
/clone="RPCI-11-481m9"			
/sex="Male"			
/cell_type="Lymphocytes"			
/clone_id="RPCI-11"			
/note="Vector: pBAC3.6; Site_1: EcoRI; Site_2: EcoRI; RPCI11 Human Male Library"			
ORIGIN			
Query Match 7.4%; Score 33.8; DB 28; Length 695;			
Best Local Similarity 49.2%; Pred.No.1.4e+02;			
Matches 89; Conservative 0; Mismatches 92; Indels 0; Gaps 0;			
Oy	275	TGACTCGAATGGTTCAGAAATATATCCACATCGACGACGTAAGCGTAAATCTCG	334
Dd	291	TAACTATACTGTATACACAAAAGAGATTATGATCAAGTGTCACTTGTAGACAGT	232
Oy	335	ATATTACTGTCCGCAATACGGCGGTATATAAGCCGGCTGTGTTAATTATGATCACTGG	394
Dd	231	ATTTTAAGTGTGTAATCAAGTGTATCAAGATGACAGCTGTATACACACATCTATCAG	172
Oy	395	TTACCGGTGTGTTAACCATGAATGACATGCAACGAACGACGCGCTAACAGATT	454
Dd	171	CCATACATGTGACACACATTTAAAGTCAACATATAGACATGACACAAAGTATACATTAT	112

OY 455 A 455
Db 111 A 111

RESULT 14

CG178280 941 bp DNA linear GSS 21-AUG-2003
LOCUS CG178280
DEFINITION PU0DF79TD_ZM_0.6_1.0_KB_Zea_mays_genomic_clone_ZM82A0641N14,
genomic survey sequence.
ACCESSION CG178280
VERSION CG178280.1 GI:34069341
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
1 (bases 1 to 941)
AUTHORS Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.

TITLE Maize Genomics Consortium
JOURNAL Unpublished (2003)
COMMENT Other GSSs: PU0DF79TB
Contact: Cathy Whitelaw
TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@cigr.org
Seq primer: 7P
Class: sheared ends.

FEATURES

source 1..941
Location/Qualifiers
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM82A0641N14"
/clone_11b="ZM_0.6_1.0_KB"
/note="Vector: pCR4-TORO; Site_1: EcoRI; 0.6-1.0 kb high
COT selected genomic DNA library"

ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 941;
Best Local Similarity 51.0%; Pred. No. 1.6e+02;
Matches 80; Conservative 0; Mismatches 77; Indels 0; Gaps 0;
OY 17 TGGAGCATTCGCGCATGCTGCTTCTGGCAGTCTGCGCTGCGCTTCCACAAAT 76
Db 236 TTGCAAGTCTACACATGAGTGTCTATATGCAACAGTCTCGCACCGGCTTCCGCGG 295
OY 77 GGGCGCGCGGCGGATATCATTAACGGCGGCGCAATAGTTCCGCCCGGAGTCAAGTTGA 136
Db 296 GGAATACCTCCGTCGTCACACACAGCCGCGCTTGAAGTTCGCGGACTGTAACCCGA 355
OY 137 GCATTTATCAGTACGTTCCGCTTAACGCTGCGCTTGC 173
Db 356 CTCTCTTCGCAACCGCTCAACGCGGACGACTCTGGC 392

RESULT 15

CNS02G0P/c 1098 bp DNA linear GSS 01-SEP-2000
LOCUS CNS02G0P/c
DEFINITION Tetradodon nigroviridis genome survey sequence PUC-ORI end of clone
138K23 of library G from Tetradodon nigroviridis, genomic survey
sequence.
ACCESSION AL196666
VERSION AL196666.1 GI:7834816
KEYWORDS GSS: genome survey sequence.
SOURCE Tetradodon nigroviridis
ORGANISM Tetradodon nigroviridis

REFERENCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetradodon.
1

Roest Crolius,H., Jallion,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizesmes,C., Wincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
Estimate of human gene number provided by genome-wide analysis
using Tetradodon nigroviridis DNA sequence
Nat. Genet. 25 (2), 235-238 (2000)
20296633
10835645

TITLE

Roest Crolius,H., Jallion,O., Dasilva,C., Ozouf-Costaz,C.,
Fizesmes,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
Saurin,W., Bernot,A. and Weissenbach,J.
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradodon nigroviridis
Genome Res. 10 (7), 939-949 (2000)
20359837
10899143

REFERENCE

3 (bases 1 to 1098)
Genoscope.

TITLE

Direct Submission

COMMENT

Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqrefgenoscope.cns.fr)
- Web : www.genoscope.cns.fr
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradodon.

FEATURES

source 1..1098
Location/Qualifiers
/organism="Tetradodon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="138K23"
/clone_11b="G"
/note="Genoscope sequence ID : C0AG138AF12SP1-end ;
PUC-ORI"

ORIGIN

Query Match 7.4%; Score 33.8; DB 29; Length 1098;
Best Local Similarity 55.8%; Pred. No. 1.7e+02;
Matches 53; Conservative 5; Mismatches 37; Indels 0; Gaps 0;
OY 19 GCAGCATTCGCGCAATGCTGCTTCTGGCAGTCTGCGCTGCGCTTCCACAAATG 78
Db 773 GAAGTCCCTGCTGAACAGGGGAGTTTGGCGTAAAGCTGCTGAATGTTCAACGAT 714
OY 79 GCGCGCGCGGTAATCATTAACGGCGCGCGCAATG 113
Db 713 AAAGCGCGCGTTTAACTTMAAGCCACGCGCATG 679

Search completed: March 16, 2004, 04:29:02
Job time : 2232.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaactttaaagtcgc.....ccacgctaacagctactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_bcg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_vl: *
14: gb_vl: *
15: em_da: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vit: *
38: em_sy: *
39: em_hlgo_hum: *
40: em_hlgo_mus: *
41: em_hlgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	404.8	88.8	456	6 144909	144909 Sequence 58
2	404.8	88.8	2067	1 SEU43280	U43280 Salmonella
3	396.8	87.0	5103	1 STRA2301	AJ002301 Salmonella
4	396.8	87.0	22411	1 AE008749	AE008749 Salmonella
5	395.2	86.7	254050	1 AL627269	AL627269 Salmonella
6	395.2	86.7	301983	1 AE016840	AE016840 Salmonella
7	389.4	85.4	1048	1 STAGFBA	AJ000514 Salmonella
8	287.2	63.0	361	6 144908	144908 Sequence 56
9	281.6	61.8	2889	1 CSP515700	AJ515700 Citrobact
10	229.2	50.3	2920	1 CFR515701	AJ515701 Citrobact
11	229	50.2	230	1 SEU53207	U53207 Salmonella
12	227.2	49.8	4680	1 ECCSGABDC	X90754 E.coli csgG
13	227.2	49.8	10346	1 AE000205	AE000205 Escherich
14	227.2	49.8	15047	1 D90741	D90741 Escherichia
15	225.6	49.5	456	6 AX814811	AX814811 Sequence
16	220.6	48.4	306358	1 AE016759	AE016759 Escherichia
17	220.6	47.7	10190	1 AE005315	AE005315 Escherich
18	217.4	47.7	10190	1 AP002554	AP002554 Escherich
19	217.4	47.7	327773	1 ESAS15702	AJ515702 Enterobac
20	188.8	41.4	2883	1 AE015131	AE015131 Shigella
21	163.2	35.8	10370	1 AE016981	AE016981 Shigella
22	163.2	35.8	292504	1 AE016981	AE016981 Shigella
23	163.2	30.7	437	1 AP237726	AP237726 Escherich
24	140.2	14.2	1212	1 EC011756	AJ131756 Escherich
25	64.6	14.1	19201	1 D90742	D90742 Escherichia
26	64.4	9.8	72	6 AX814798	AX814798 Sequence
27	44.8	9.8	2000	6 AX655393	AX655393 Sequence
28	44.8	8.6	956	8 AK058943	AK058943 Oryza sat
29	39.4	8.6	2720	8 AK102629	AK102629 Oryza sat
30	39.4	8.6	2720	8 AP004339	AP004339 Oryza sat
31	39.4	8.6	15248	8 AE001012	AE001012 Archaeogl
32	38.8	8.5	13264	1 AF3429315	AF3429315 Homo sapi
33	38.2	8.4	125020	9 AF343445	AF343445 Lactobaci
34	36.2	7.9	1425	1 AC116977	AC116977 Dictyoste
35	36.2	7.9	302156	3 AC116977	AC116977 Pan trogl
36	36.2	7.9	158267	9 AL359644	AL359644 Human DNA
37	35.8	7.9	172600	2 AC011969	AC011969 Homo sapi
38	35.8	7.8	193168	2 AE016786	AE016786 Pseudomon
39	35.6	7.8	301214	1 AC104285	AC104285 Oryza sat
40	35	7.7	117541	2 AY247273	AY247273 Medicago
41	34.4	7.5	560	8 HAAR165	MS9123 Haemaphysob
42	34.4	7.5	1553	1 AX416572	AX416572 Sequence
43	34.4	7.5	2751	6 AC146008	AC146008 Pan trogl
44	34.4	7.5	151828	2 AC130549	AC130549 Mus muscu
45	34.4	7.5	160449	10	

ALIGNMENTS

RESULT 1
LOCUS 144909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen. and Clouthier,S.C.,
TITLE Methods and compositions comprising the agfa gene for detection of
Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

Pred. No. is the number of results predicted by chance to have a

FEATURES
Source 1: 456
Location/Qualifiers
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 88.8%; Score 404.8; DB 6; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.9e-106;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGGCGCATTTGCGAGCAATGATGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTAAAGGCGCATTTGCGAGCAATGATGTTCTGGCAGTCTGGCT 60
QY 61 GCGCTGTTCCACATGCGCGCGCGGCGGATTCATACGCGCGCAATAGTTCGCG 120
DB 61 GCGCTGTTCCACATGCGCGCGCGGCGGATTCATACGCGCGCAATAGTTCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTTACATGACGTTCCGCTAACGCTGCGCTTGCCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTTACATGACGTTCCGCTAACGCTGCGCTTGCCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACCATTTCCGAGCGGTTATGCTAAACGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACCATTTCCGAGCGGTTATGCTAAACGCGCGAT 240
QY 241 GTAGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCGCAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCGACCATGACCAAGTGGACGCTAAATAATCCGCGATTTCTGCGCAATATGATCAG 360
DB 301 GCGACCATGACCAAGTGGACGCTAAATAATCCGCGATTTCTGCGCAATATGATCAG 360
QY 361 CTGTTACCGCGTGTGTTACCCATGAATGACCATGACCATGACCATGACCATGACCAT 420
DB 361 AATAACGCGCGCGTGTGTTACCCATGAATGACCATGACCATGACCATGACCATGACCAT 420
QY 421 GGTTTGGCAACAACGCGCGGCTAACAGTATTTAA 456
DB 421 GGTTTGGCAACAACGCGCGGCTAACAGTATTTAA 456

RESULT 2
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbria-like protein
precursor (agfB), thin aggregative fimbriae precursor (agfA), and
AgfC (agfC) genes, complete cds.
VERSION U43280.1 GI:1184712
KEYWORDS
SOURCE
ORGANISM *Salmonella enteritidis*
Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.
1 (bases 1 to 2067)
Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, E.C.,
Munro, C.K., Kay, C.M., Baner, P.A., Peterlin, P.I., and Kay, W.W.,
DNA-based diagnostic tests for *Salmonella* species targeting agfA,
the structural gene for thin, aggregative fimbriae
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
94013373
8104955
2 (bases 1 to 2067)
Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
Kay, W.W.
TITLE *Salmonella enteritidis* agfBAC operon encoding thin, aggregative
fimbriae
J. Bacteriol. 178 (3), 662-667 (1996)
MEDLINE 96146512
PUBMED 8550497
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
3 (bases 1 to 2067)

AUTHORS Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
Victoria, BC V8W 3P6, Canada
Location/Qualifiers
1: 2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKNLKLPMLTILGAPGATATATYDLSREYFAVNLKSPFN
SISQSAVGNASAIILQKSGNKANITVGTQKTAIVVQKSHAIRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids;
fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/transl_table=1
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLKVAPAIIVSGSALAGVYPQWGGGNNNGSGSPDST
LSIYQGSANALALQSDARKEETITGCGAGADVGAGADNSTIELTONGFRNAT
IDQWNAKSDITVQGYGNNALVNTASDSVVRQVGGNNATANQY"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbria"
/note="thin aggregative fimbria subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=1
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTLLILAAASNOITFTTQGGDIYVIVPVLNBPVCVOYITL
SYRDVGCGSHRQKQTLSPANOPIELSRLSVNISSDSKIVTIVSDQSLHLSQO
WPSAQ"

ORIGIN

Query Match 88.8%; Score 404.8; DB 1; Length 2067;
Best Local Similarity 93.0%; Pred. No. 1e-105;

gene 3344. .3376
/gene="csGBA"
-35_signal 3344. .3349
/gene="csGBA"
-10_signal 3371. .3376
/gene="csGBA"
gene 3469. .3924
/gene="csGBA"
CDS 3469. .3924
/gene="csGBA"
/function="nucleator"
/codon_start=1
evidence=experimental
/transl_table=1
/product="CesB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/transl_table=1
/translation="MKKLLFMVLTITGPGIATVNTDILARSEYFVAVNELSKSFN
QAAITGVGTDNARVROSGKSLVITSGEGNNRAVDGANGYNFAYIEOTGNANDA
SISQSAAGNSAAIIQKSGNKANITQYGTQKTAIVVQKSHMAIRYQF"
sig_peptide 3469. .3531
/gene="csGB"
mat_peptide 3532. .3921
/gene="csGB"
/product="CesB protein"
/function="nucleator"
gene 3966. .4421
/gene="csGA"
CDS 3966. .4421
/gene="csGA"
/function="fiber subunit"
/codon_start=1
evidence=experimental
/transl_table=1
/product="CesA protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/transl_table=1
/translation="MKLKVAPAIIVSGSALAGVFWQWGGGNGHGGNNGSPST
LSDYQGSAAALALQSDARKEITTSQNGADVGQADNSTIETLNGRRNAT
IDWNAKNSDITVQYGSNNALVNOTASDSVWVRQVGFGNATANQY"
sig_peptide 3966. .4025
/gene="csGA"
mat_peptide 4026. .4418
/gene="csGA"
/product="CesA protein"
/function="fiber subunit"
gene 4483. .4809
/gene="csGC"
CDS 4483. .4809
/gene="csGC"
/codon_start=1
/transl_table=1
/product="CesC protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translation="MHTLLIALASKQITFTTQGGDIYVYIPQVTLNPEVCQVOTL
SVRDVGQSGHQQKOTLPLANQPIELSRSLVNSHSDSVKIIIVTSDDQSILHSQQ
WPSSAQ"
sig_peptide 4483. .4533
/gene="csGC"
mat_peptide 4534. .4806
/gene="csGC"
/product="CesC protein"
gene 4940. .5103
/gene="ORF103"
CDS 4940. .5103
/gene="ORF103"
/codon_start=1

ORIGIN
/transl_table=1
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/db_xref="SWISS-PROT:O54295"
/translation="MSVKKNIIPALIGLCIAFPIHSAVQGVQVGHFRGATVEPL
CDISTHANI"

Query Match 87.0% Score 396.8; DB 1; Length 5103;
Best Local Similarity 91.9%; Pred. No. 2,3e-103;
Matches 419; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCAGCATTCGACGATCGTGTCTTCGACAGTCTGCTGCT 60
Db 3966 ATGAACCTTTTAAAGTGCAGCATTCGACGATCGTGTCTTCGACAGTCTGCTGCT 60
QY 61 GCGCTCGTTCACAAATGGGCGGCGGCGGTATCTATTAACGCGCGGCAATAGTTCGGC 120
Db 4026 GCGCTCGTTCACAAATGGGCGGCGGCGGTATCTATTAACGCGCGGCAATAGTTCGGC 120
QY 121 CCGGACTCAACGTTGACATTTATAGTACGTTCCGCTTACGCTGCTCTTGCAT 180
Db 4086 CCGGACTCAACGTTGACATTTATAGTACGTTCCGCTTACGCTGCTCTTGCAT 180
QY 181 AGCGATGCCGTTAAATCTGAACGACATTCACGACGCGGTTATGTAACGCGCGCAT 4145
Db 4146 AGCGATGCCGTTAAATCTGAACGACATTCACGACGCGGTTATGTAACGCGCGCAT 4145
QY 241 GTAGCCAGGCGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 4205
Db 4206 GTAGCCAGGCGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 4205
QY 301 GCCACATTCACGATGGAACGCTTAAACTCCGATATACGTGCGCAATATGATCAG 4265
Db 4266 GCCACATTCACGATGGAACGCTTAAACTCCGATATACGTGCGCAATATGATCAG 4265
QY 361 CTGGTACCCGCTGTTGTTCCATGAAATGACATGCAATGCAATGCAATGCAATGCAAT 4325
Db 4326 AATGACCCGCTGTTGTTCCATGAAATGACATGCAATGCAATGCAATGCAATGCAAT 4325
QY 421 GGTGGTGGCAACACCCGCTTAAACGATTAACGATTAACGATTAACGATTAACGATTAAC 4385
Db 4386 GGTGGTGGCAACACCCGCTTAAACGATTAACGATTAACGATTAACGATTAACGATTAAC 4385

RESULT 4
AE008749 22411 bp DNA linear BCT 23-APR-2003
LOCUS AE008749
DEFINITION Salmonella typhimurium LT2, section 53 of 220 of the complete genome.
ACCESSION AE008749 AE006468
VERSION AE008749.1 GI:16419641
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
CONSTRM
TITLES
JOURNAL
The Salmonella typhimurium Genome Sequencing Project
Submitted (29-MAR-2001) Genome Sequencing Center, Department of

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI33283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Julio Collado-Vides and
RegunoldB;
http://kinich.cifn.unam.mx:8850/db/regunoldb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES
source location/Qualifiers

1..22411
/organism="Salmonella typhimurium LT2"
/mol_type="genomic DNA"
/strain="LT2; SGSC 1412; ATCC 700720"
/db_xref="ATCC:700720"
/db_xref="taxon:99287"
/note="LT2"
434..1308
/gene="phoH"
/note="synonym: STM1126"
434..439
/gene="phoH"
/note="putative RBS for phoH; RegulonDB:STMS1H001398"
454..1308
/note="putative RBS for phoH; RegulonDB:STMS1H001398"
/gene="phoH"
/note="similar to E. coli Phob-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"
/codon_start=1
/transl_table=1
/product="Phob-dependent ATP-binding pho regulon
component"
/protein_id="AAL20057.1"
/db_xref="GI:16419642"
/translation="WCANRVQENAGDVYIQLKLSNGRQAKVIKAREAKRLARDS
RSHKQREESVTSLVOMGGVYALIGMARSDRTSVKANEQAHYLNAIDSKOILPAT
GEACGGTWISAAKAAALIHKIDVERIIVTPVLAQADELDGLFDIAEKAPYFRPV
YDVILKRLGASFMQYCLREIPEIKVEIAPFAVGRGTENAVVILDEAONKVTAAOMKRF
LTRIGENVTVVNDITQCCLPRGVRSGSLDALRFEDEWVGIVHNKDDCVASALC
ORTIHAYS"
complement(1414..2303)
/gene="STM1127"
complement(1414..2295)
/gene="STM1127"
/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%
identity in aa 15 - 264"
/codon_start=1
/transl_table=1
/product="putative transcriptional regulator"
/protein_id="AAL20058.1"
/db_xref="GI:16419643"
/translation="MEQPPRLKPKGLIDTIGAMQKSTTRASQRIAGYIAPQVTO
SSINDSRETQAGEATVIRFRTLGYGQDFQKNDLIELATTSDDSSPLDVEE
SDDHAIGLKQNTISNVLSTNLIDMOQLGVADLRHCHSYVIFGVSSGTTAD
MKHLMRIGLRGDAVSNHFMWQATLLKAGDVAMGVSHSGTSPETVHSIRLAQACA

TTVAITNIGSPICEADFCILINRGQMLQGDSIGTKAQLFVFDLLYTLVQSSPE
QARSKIRTNALDMTK"
complement(2298..2303)
/gene="STM1127"
/note="putative RBS for STM1127; RegulonDB:STMS1H001399"
complement(2580..4088)
/gene="STM1128"
complement(2580..4076)
/gene="STM1128"
/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%
identity in aa 7 - 478"
/codon_start=1
/transl_table=1
/product="putative sodium/glucose cotransporter"
/protein_id="AAL20059.1"
/db_xref="GI:16419644"
/translation="MITHSRGIVNYLVLFEGYILAMLVGVYSSRROKTDADYPRGGR
VPGWAGVSPATLTSITPMSIPAKFTSDWPTIIGQYLAIALIPLVFPYIPERK
IKVTSAYEYLEARDVRCRLPASNSPMLPHIGRAIITFTLVALLRPIAIDPILVL
LISVWCILITVMGGISEGVITVDVIOGLLSALILFIVLCIKVQGIIDEFTVQQA
DKFPPATQFMSTWSTSTVPLMIGLFPANTIQOFPASQDVQRYVTDISEBTKTLT
NAKLAVIIPVEFPAIGSALFVYQOHPQLPAGTNGIILPLFVTEMPVIGAGLIA
AIFPAQSSISSINSISCFNSDIYORLSHKRTPENRMKILAKVILVAGLSSAAS
VWLVMADSRIMDAFNSLIGMGPMGTGLMIGIFFRANAGSAVIGIISVITVLCG
RYATDLNFFPYGVGSLSVVISGYIFRPLPAPAPPLTLDKPEPKTL"
complement(4083..4088)
/gene="STM1128"
/note="putative RBS for STM1128; RegulonDB:STMS1H001400"
complement(4413..5099)
/gene="STM1129"
complement(4413..5093)
/gene="STM1129"
/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20060.1"
/db_xref="GI:16419645"
/translation="MSILARLEQSHENGLIVSCQPVGSPMDKPRIVAAQAAS
AGAVAVIEGENTRTVPHLSVPIIGIRDLGSGRITPYQDDVALAQADII
APASRFSRVDLSILTRIRHGLTAMADCSYNEISCHQKIEIRITLSYGTQ
ITPVEDPLAVTQLSHNGCRVIAEGRYNTPALANALIEHMAVTVSALIRISHICQ
WFSHAVKR"
complement(5094..5099)
/gene="STM1129"
/note="putative RBS for STM1129; RegulonDB:STMS1H001401"
5599..6759
/gene="STM1130"
5599..6759
/gene="STM1130"
/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%
identity in aa 40 - 403"
/codon_start=1
/transl_table=1
/product="putative inner membrane protein"
/protein_id="AAL20061.1"
/db_xref="GI:16419646"
/translation="MGQMKNPKKMTLALCSVAITTSQYATLLPDIDPELKNCG
AIDNGVITVGLGTAGTSYTKIDKQKHWEIISPGGARGQSSVFLNDELITVVG
GVGRKNSSEPLQVYSDVYKTSVPKNTQKVDIISPVGLTGHGKLSNETVLITGAVN
EHI.FDKYFIDIAAADADESEKNKVIYVFKPAKDDVFNKIVIPYLNKENTWKAAGEL
PCAGTIGSSSVMGNNFLMLINGELKPLRFVDVLYRAMMDKLTMLKNSQLPSPGCG
QOGLGAFSGYSHGVLLVGGANFPKAPKQNTYNGKXSHSGINKKRDVYGLINCH
WQMGKMKQBLIGVSVSYGDVFLIGERAKKAPSVSVTFMRDNLILIK"
6793..7497
/gene="STM1131"
6793..6798
/gene="STM1131"
/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

```
6805. .7497
/gene="STM1131
```

(AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 26% identity in aa 4 - 239"

```

/product="putative outer membrane protein"
/protein_id="AAL20062.1"
/db_xref="GI:16410647"

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

```

RBS      /gene="STM1132"
          7768.  .7773
          /gene="STM1132"
          /note="putative RBS for STM1132; RegulonDB: STMS1H001403"
CDS       7780.  9060
          /gene="STM1132"

```

```
(AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 57%
identity in aa 25 - 418"
/codon_start=1
/translation_table=1
/product="putative sugar transport protein"
```

Query Match	87.0%;	Score 396.8;	DB 1;	Length 22411;
Best Local Similarity	91.9%;	Pred. No. 2.6e-103;		
Matches 419;	Conservative			

6

3

to AMPN_ECOLI (869 aa), 94% identity in 869 aa overlap" /codon_start=1

```
/transl_table=11
/product="aminopeptidase N
```

```

/protein_id="CAD08183.1"
/db_xref="GT:16502323"

```

```

/db_xref="GOA:Q8Z7T0"
/db_xref="EBI:Q8Z7T0"

```

```

/translation="MTQQPQAKYRHDY
HGARDARDPDR"

```

NSAFDAFLKLDGEDLTIVSIHVNDAPW
NTALEGLYQSGDALCTQCEAEGFRHIT
DINOCIT

KVAQGELENGRHWVQWQDPFPKPCYLF
NLDRAFWAMTSLKNSMKWDETREGI,EY

LARTDTATDKDYLDIERVIGHEYFHNW
GSAVRNRSNVRTMRGLQFAEDASBMA

TLGEENFQKGMQLYFERHDSAAATCD
KDDYNPETEOYTI.TISORTBATDORF

VLNVTQAEQTFTEFDNVYFQPVPA LCE
AAOSTLATYTKIMHAPROGSA

IAELFEVIDPIAIAQVREALTRTLAEI
ACIPEI TECE

QDGLVMDKWEILQSTSPAENVLETVRSI

AQUGSGYQFLVEMLTDLNSRNPQVASRI
LSGDLYEKITKALA"

```
532. .1332
/gene="STY1078"
```

```
/note="Pfam match to entry P
family M1 score 345.107
```

1060. .1089
/gene-llctm1.070"

Page 6

/note="PS00142 Neutral zinc metalloproteinases,
zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit
to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82789"
/db_xref="Swiss-Prot:Q82789"
/translation="MYPPFVKALFQDLPREAHFTFOQLRRITGTPLEALVROKVP
KPTVMGILFKNPLGLAAGLDKDECIDALGAMGSLGISTVTPRPQNDKRLPR
LVDAEGLINRGFNGLGVNLENVKKAFPGIIGINIGKKOTPEVNGKDYLCME
KVYVAGYIAINISPMTPGRLTYOYGDALDLDLTALIKKNDLOVHHKVPVAVY
APDCERELIOVADSLRNIDVITATTTTDRSLVQMKRCQQTGGLSGRPLKST
EIRRLDELKQLPIIGVGIDISVIAAREKIAAGATLVQISGIFKGPPLKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase,
Dihydroorotate dehydrogenase, score 563.50, E-value
1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to
YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note
discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTREMBL:Q82758"
/translation="MMRIKPDNMWYDEEDHDMMLDLANGMLFRSFRSKMLTPDA
FCTGFCVDALVYSEFEKCRDELTKEORALVINALVALIRYLKPOMKSMHFAV
GEMWTGCTGAASVMSLDTAEQVNLVVERENALCLLAPGVIVAGRTMGLDALK
IMNDRLKPVCHSFLSQAV"
complement(4716..5825)
/gene="STY1081"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863
(369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82787"
/db_xref="SPTREMBL:Q82757"
/translation="MATLSRLFIHPVSMKGIQTLHALADI SGIAFDRIMVWESDGT
FITAPQPMWRFTPSPLHDGLHLTAPDSSALVRFDPDPTVMGNHFAVA
PTAINOMLSEFSSVDOLCVWGPOLTRVRKHNAVPLGDFDGYLLTNEASRLDRA
RCAGVQMEQFERNLVSGVAWEEDSMKYLRTGDPVFDVVKPESRITPTVSEKCO
KH8SGEPLATLQARTNODNGVDNPGNLLARNSGVRVGEDEEITLTAAPAKAGATT
LDGSEVPEKHPKDSVITDQSGQTFEGNNQVLLLEQLNNGIRLIPYSGRAGICGCRIR
LLEGEVSPKRSAGWDGDTILSCSVCKTALRLEN"
complement(4737..4955)
/gene="STY1081"

/note="Pfam match to entry PF00111 fer2, 2Fe-2S
iron-sulfur cluster binding domains, score 55.90, E-value
8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical
protein h10116 H10116 SW:YCBY_HAEIN (P44524; P43945) Faeta
scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI
(702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82786"
/db_xref="SPTREMBL:Q82786"
/translation="MNSLPASTARGLEELIKELKAGVCCOVVQGVHRCQTRLI
YOSLMSRLSRILPEWCKVSDLDLYFGVQALNTEIFNPGATVAVHPSGLNDTI
RNSOYGMKVKDALVDAFTKRLPRPVNDESDPLRLNWLNKETASIALDLSGLH
LRGYRDRTGAPIKETLAAAIIVMSGQPTPLDPMCGSGTLLIEAAMATDRAPIGL
HRGHGFGSGMAOHDEAIWQRYKAEOTRAKGLAEYSHFSGSDSARVERASNR
RAGIGELITPEVDVQALSNPLPKPGVTIVSNPVGPERLSDSPALTAISLIGRTMK
NORGKNTLSFSA5PDLGSLQLRADQFAKNGPLDVCYKNTIATITADSKPATYA
EDIANRLRKLKLEKMAROEICVRLYADLPEVNAADRQDMVAVI0YAPKTV
DAQKARQLPDIILAAITL5VIGIPENKVLVITREOKKNOYOKXSEGEFLEYSYNA
RLWNLITDYLDITGLFDHRIARMLAGMSGXPLNFSTYGSASVAGKANTTT
VDMSTRYLEWAEERLRLNGSGRAHRLI0ADCGMLLEANEOPDLIPIDPTFENSRR
MEESPVDYRDLVHMDLRLKRLKKGITIMSNKRGFRMDLEGLAEILTA0EITQKT
LSPDFARNRQIHNKMLIRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020,
Uncharacterized protein family UPF0020, score 781.30,
E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020
signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in
526 aa overlap
Faeta hit to YJUX_ECOLI (554 aa), 34% identity in 524 aa
overlap

Query Match 86.7%; Score 395.2; DB 1; Length 254050;
Best Local Similarity 91.7%; Pred. No. 9,6e-103;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY	1	ATGAACCTTTAAAGTGGAGCATTCGACGACATGCTAGTTCTGACAGTCTGGCT	60
Db	88906	ATGAACCTTTAAAGTGGAGCATTCGACGACATGCTAGTTCTGACAGTCTGGCT	88965
QY	61	GGCGTCGTTCCACAAATGGGCGGGCGGTATCATATACGCGCGCAATAGTTCCGGC	120
Db	88966	GGCGTCGTTCCACAAATGGGCGGGCGGTATCATATACGCGCGCAATAGTTCCGGC	89025
QY	121	CCGACCTCAACGTTGACATTTATCAATACGTTCCGCTAACGCTGGCTTCTGCGAA	180
Db	89026	CCGACCTCAACGTTGACATTTATCAATACGTTCCGCTAACGCTGGCTTCTGCGAA	89085
QY	181	AGGATGCGCGTAAATCGAAAGACCATTAACCAAGCGGTTATGTTAAGCGGCGCAT	240
Db	89086	AGGATGCGCGTAAATCGAAAGACCATTAACCAAGCGGTTATGTTAAGCGGCGCAT	89145
QY	241	GTAAGCCAGGCGTGAATAGTACTATTGAACTGACAGATGTTTACAGAAATAT	300
Db	89146	GTAAGCCAGGCGTGAATAGTACTATTGAACTGACAGATGTTTACAGAAATAT	89205

gene
RATNALGATPMDLMSILQSMOAGCLRAKLEIT"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
complement (5597..6295)
/gene="flgH"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKYALHAPYVMAIMVATLTGCAWTPAKPIVQAGTTPAQPIPGV
PVANGSIPOSQPINGYQPLFEDRPENIDTLTIVQENVASKSSAASRDGKT
SPGFDTPRYLQGLFGNSRADMEASGGSFSGKGAASNTFSGTLTATVQVLANGN
LHVGEKOIAINQCTEFIRFSGVNPRTISGNSVSPQVADARIEVYNGYINEAQN
MGMLORFFLNLSPM"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
complement (6350..7132)
/gene="flgG"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISLWIAKTGLDAQQTNDVIANNLAVSTNGFKRORAVFEDL
LYQTRPGASSEQTLPGLQIGTVRPATERLSQGLSQTNNSKVAIKGQGF
FOVNLPGTSAYTRDGSFOVQNGQLYAGFOVPAITIPANLISITIGDVSAT
QQCGAARVQVQLTLTFMNDTGLSISGENIETOSGAPNESTPGINGAGLLYGY
VETSNVAEELVMIQVORAYEINSKAVSTTDMLQKLTQL"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
complement (7146..7901)
/gene="flgF"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHAIYAMGASQTLNQCAVITNSLNASTPGRQALNLRAY
PVDELATRLTLVASTPGADMTQGLDYTSRPLDVALQODGMVVOAAGAEQYTN
GNIOVPTGQLIQGHFVIGEGEITVPEGSEITADGTSALNPDDPNTVAVPR
LKLYKAGNEVQRSDGGLFRLTAEOAERGAVLADSRIMSVLEGSNVKPYEAMT
DMINARPEMQKVITSVDENGRANQLLSMS"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
complement (7922..9133)
/gene="flgE"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSEFSQAVSGMAATNIDVGNRIANSATYGFKSGASPADMA
GSKVGLGVKAGITQDFTDGTITNGDLVALSONGFRLVDSNGSVFYSRNGQFCL

gene
86.7%; Score 395.2; DB 1; Length 301983;
Best Local Similarity 91.7%; Pred. No. 9.8e-103;

Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;
QY 1 ATGAACTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTGTGGCAGTGTCTGCT 60
Db 37210 ATGAACTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTGTGGCAGTGTCTGCT 37251
QY 61 GGGCGTCTTCAAAATGGGGGGGGGGGGTATATATAACGGCGGGGCAATAGTTCCGGC 120
Db 37250 GGGCGTCTTCAAAATGGGGGGGGGGGGTATATATAACGGCGGGGCAATAGTTCCGGC 37191
QY 121 CCGGACTCAACGTTGAGCATTTATAGTACGTTCCGCTAAGCGTGGCTGCTCTGCA 180
Db 37190 CCGGATTTCCAGCTTACGATTTATAGTACGTTCCGCTAAGCGTGGCTGCTCTGCA 37131
QY 181 AGCGATCCCGGTAATCTGAAACGACCATTAACGAGCGGTTATGTTAGTACGCGCGCAT 240
Db 37130 AGCGATCCCGGTAATCTGAAACGACCATTAACGAGCGGTTATGTTAGTACGCGCGCAT 37071
QY 241 GTAGCCAGGGTGGGATTAATAGTACTATGAACTGACTGAGATGGTTTCAGAAATAT 300
Db 37070 GTAGCCAGGGTGGGATTAATAGTACTATGAACTGACTGAGATGGTTTCAGAAATAT 37011
QY 301 GCCACCATGACACGATGAAAGCTAAAACTCCGATATTACTGCGCCATATATGATGAC 360
Db 37010 GCCACCATGACACGATGAAAGCTAAAACTCCGATATTACTGCGCCATATATGATGAC 36951
QY 361 CTGGTACCCGTTGTTTACCATGAAATGGACATGCAAGCGTATATGTCGTCAGGTT 420
Db 36950 AATTAACCCCGCTGCTGTTATTCAGACCGCATCTGATTCAGACGTTATATGTCGTCAGGTT 36891
QY 421 GGTTCGGCAACAGCGCAGCGCTAACCATATTTAA 456
Db 36890 GGTTCGGCAACAGCGCAGCGCTAACCATATTTAA 36855

RESULT 7
STRAGBA 1048 bp DNA linear BCT 26-JAN-1998
LOCUS Salmonella typhimurium agfB and agfA genes.
DEFINITION
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1
Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE 2 (bases 1 to 1048)
Sukupolvi,S.S.
TITLE Direct Submision
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllynkatu, 20520, FINLAND
FEATURES
source
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CA04150.1"

Query Match	85.4%;	Score 389.4;	DB 1;	Length 1048;
Best Local Similarity	91.0%;	Pred. No. 2.7e-101;		
Matches 414;	Conservative	0;	Mismatches 41	

RESULT 8
T44000

Query Match	63.0%;	Score 287.2;	DB 6;	Length 361;
Best Local Similarity	99.0%;	Pred. No. 8,4e-72;		
Matches 289;	Conservative	0;	Mismatches	7

Db


```

/translation="MFNEFHSIHGHTLLITKPSLOATALLQHLKOSLALSGKLTNIO
RSLDSSSCIVLDMMEADKLIHYQDNLSRKNNIKTLLIPDPYPRDIENMP
HINCFYVADDEBVRNMGLOGLIRECPFSOKLASYLTHSGNRYNSSEALLTHRE
KEILANKRIGASNEIARSFLISENTVKTLYNLFKLIAYNRTQAVSWANDILRR"
1623..2078
/gene="csgb"
1623..2078
/cds
/function="nucleator"
/codon_start=1
/translation="nucleation component of curlin monomers"
/protein_id="CAD56673.1"
/db_xref="GI:31790493"
/translation="MKRKLPMMLTVLGPAGIASNTSYDLANSEYPAVNELSKSP
QALIGOVGNNSAKRQSKLSVSOEGSGNRAKVDGCAVPAIAGSHSND
SISOSYGNVTAMIIQKSGNKNITQYQKTAVVVQROSMAIRVQR"
2119..2571
/gene="csga"
2119..2571
/cds
/gene="csga"
/function="curli subunit"
/codon_start=1
/translation="curlin-csga protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLIQVAFPAIIVSGSALAGSPQWGGGGGGSSGPESTL
SIYOSGVNNAALQSDARKSDVTIHQGFNGNAGVGGSDNSTIDLQNFKNATTI
DQWNGKNSDITVSGYGHNAALVNGTASDSVLVHQVFGNNAATANGY"

ORIGIN
Query Match 61.8%; Score 281.6; DB 1; Length 2889;
Best Local Similarity 77.6%; Pred. No. 4.3e-70;
Matches 354; Conservative 0; Mismatches 99; Indels 3; Gaps 1;

Qy 1 ATGAAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTTCTGGCAGTCTGGCT 60
Db 2119 ATGAAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTTCTGGCAGTCTGGCT 2178
Qy 61 GGCCTCGTTCCAAATGGGGCGCGCGGTATCATMACGGCGCGGCAATAGTTCCGGC 120
Db 2179 GGTTCGTGTTCCGAATGGGGCGCGCGG---TGGCGGCGGCGGAGCAGCTCCGGC 2235
Qy 121 CCGGACTCAAGCTTGAAGATTTATGATGAGGTTCCGCTAACGCTGCTCTGTGCA 180
Db 2236 CCGGAAATGACCTGAGCATTTATGATGAGGATCAATTAACGCGCTTGCCTGCA 2295
Qy 181 AGCGATGCCGTAATCTGAAGACGACATTAACCGAGCGGTTATGTAACGCGCGCAT 240
Db 2296 AGCGATGCCGTAATCTGAATGACGACATTAACCGAGCTTTGTAACGCGCGAGC 2355
Qy 241 GTAGGCCAGGATGCGGATTAATGATTAAGTCAAGATGATGTTTCAAAATAT 300
Db 2356 GTAGGCCAGGATGCGGATTAAGTCAAGATGATGTTTCAAAATGATGTTTCA 2415
Qy 301 GCCACCATGACCAATGGAAGCTTAAATCTCCATTTATCTGTCGCCCAATATGATAG 360
Db 2416 GCCACCATGACCAATGGAAGCTTAAATCTCCATTTATCTGTCGCCCAATATG 2475
Qy 361 CTGGTTACCGGTGTTGTTACCATGAATGGAACATGCAACGTAATGTCGTCAGGTT 420
Db 2476 CATACGCCGACAGTGTGAACAGACTGCTCCGATTCAGAGCTTCTGTGATCAGGTT 2535
Qy 421 GGTITTTGGCAACAACGCGCGCTTAACGATATTA 456
Db 2536 GGTITTTGGCAACAACGCGCGCTTAACGATATTA 2571

RESULT 10
CFR515701 2920 bp DNA linear BCT 24-JUN-2003
LOCUS Ctrbacter freundii csgb gene, csga gene and csgd gene.
DEFINITION

```

```

ACCESSION AU515701
VERSION AU515701.1 GI:31790495
KEYWORDS csga gene; csgb gene; csgd gene; curlin-csga protein; nucleation
SOURCE component of curlin monomers; regulatory protein.
ORGANISM Ctrbacter freundii
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Ctrbacter.
REFERENCE 1
AUTHORS Zoga J, X., Bokranz, W., Nimitz, M. and Romling, U.
TITLE Production of Cellulose and Curli Fimbriae by Members of the Family
REFERENCE Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
AUTHORS Infect. Immun. 72 (7), 4151-4158 (2003)
TITLE 2 (bases 1 to 2920)
JOURNAL Romling, U.
Direct Submission
Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
location/qualifiers
SOURCE 1..2920
/organism="Ctrbacter freundii"
/mol_type="genomic DNA"
/isolate="fec4"
/isolation_source="faeces"
/species="Homo sapiens"
/db_xref="taxon:56"
/country="Germany"
/complement(211..861)
/gene="csgd"
complement(211..861)
/function="necessary for transcription of the csgrA
operon"
/codon_start=1
/translation="putative regulatory protein"
/protein_id="CAD56673.1"
/db_xref="GI:31790496"
/translation="MFNEFHSIHGHTLLITKPSLOATALLQHLKOSLALSGKLTNIO
RSLDSSSCIVLDMMEADKLIHYQDNLSRKNNIKTLLIPDPYPRDIENMP
HINCFYVADDEBVRNMGLOGLIRECPFSOKLASYLTHSGNRYNSSEALLTHRE
KEILANKRIGASNEIARSFLISENTVKTLYNLFKLIAYNRTQAVSWANDILRR"
1624..2082
/gene="csgb"
1624..2082
/cds
/gene="csgb"
/function="nucleator"
/codon_start=1
/translation="nucleation component of curlin monomers"
/protein_id="CAD56674.1"
/db_xref="GI:31790497"
/translation="MKRKLPMMLTVLGPAGIASNTSYDLANSEYPAVNELSKSP
QALIGOVGNNSAKRQSKLSVSOEGSGNRAKVDGCAVPAIAGSHSND
SISOSYGNVTAMIIQKSGNKNITQYQKTAVVVQROSMAIRVQR"
2123..2572
/gene="csga"
2123..2572
/cds
/gene="csga"
/function="curli subunit"
/codon_start=1
/translation="curlin-csga protein"
/protein_id="CAD56675.1"
/db_xref="GI:31790498"
/translation="MKLIQVAFPAIIVSGSALAGSPQWGGGGGGSSGPESTL
SIYOSGVNNAALQSDARKSDVTIHQGFNGNAGVGGSDNSTIDLQNFKNATTI
DQWNGKNSDITVSGYGHNAALVNGTASDSVLVHQVFGNNAATANGY"
ORIGIN
Query Match 50.3%; Score 229.2; DB 1; Length 2920;
Best Local Similarity 70.6%; Pred. No. 5.9e-55;
Matches 322; Conservative 0; Mismatches 128; Indels 6; Gaps 1;

```

QY 1 ATGAACTTTTAAAGGCGAGCACTTGGCAGCAATGTAATTTCTGGAGTGGCT 60
 Db 2123 ATGAACTTTTAAAGGCGAGCACTTGGCAGCAATGTAATTTCTGGAGTGGCT 218
 QY 61 GCGCTGCTTCCAAATGGGCGGCGGGGTAATCATACGGGCGGCGCAATGTTCCGGC 120
 Db 2183 GGTTGTTCCGCAATGGGCGGC-----ATCATCATGTGTGTGGCAGTAATTATGCG 223
 QY 121 CCGAGCTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGTTGCTCTGCAA 180
 Db 2237 CCAAGCTCTTCACTGAGTATCTTCACAATATGGGTCAAAACACTTCAAAATGCTGCAA 229
 QY 181 AGCGATGCCCGTAAATCTGAAACGACCATTAACCAAGGGTTATGTGTAACGGCGCGAT 240
 Db 2297 AGTGATGGCGGTAAATCATGATGTCACATCAACAACACGGGCTGGCAACGAGCTGTT 235
 QY 241 GTAGGCGAAGGTCGGATTAATGTAATTAATTAATGTAATGTAATGTTTCAGAAATAAT 300
 Db 2357 GTTGGCCAGGCGGTCTATACAGTACCATTAACCTGAAACAGACCGGCTTCCAGAAAGT 2416
 QY 301 GCCACCATTGACACAGTGAAGCGTAAAAATCTCGAATTAATTAATCTGGGCCAATATATCAG 360
 Db 2417 GCCATATGATCAGTGAATGCAAAAAATGCTGAATTAATGCTGAATCCCAAGTTCGGTGC 2478
 QY 361 CTGCTTACCCCGTGTGTTAACCATGAATGGCACAATGGACGTAATGTGTCGTGAGGTT 420
 Db 2477 CGCAACGTCGTTGTTAATAGACAGGCTCTGACTCCAAATGTGCTATTTCAGAGGTT 2536
 QY 421 GGTTTTGCACACAGCCACGGCTAACAGATTTAA 456
 Db 2537 GGCTTTGTAACACGCTAACGCTTAAACAACCTTA 2572

FEATURES	Source
LOCUS	SEU53207
DEFINITION	Salmonella enteritidis SEF17 fimbria (agfa) gene, partial cds.
ACCESSION	U53207
VERSION	U53207.1
KEYWORDS	GI:1293677
ORGANISM	Salmonella enteritidis Salmonella enteritidis Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.
REFERENCE	1 (bases 1 to 230) Cox, J.M., Egglezos, S. and Woolcock, J.B. Virulence of Salmonella enteritidis in chickens correlates with colony morphology and expression of SEF17 fimbriae
AUTHORS	Unpublished
TITLE	2 (bases 1 to 230) Cox, J.M., Egglezos, S. and Woolcock, J.B.
JOURNAL	Direct Submission
REFERENCE	Submitted (01-APR-1996) J.M. Cox, Food Science and Technology, The University of New South Wales, Sydney, NSW 2052, Australia
AUTHORS	Location/Qualifiers
JOURNAL	1. 230
REFERENCE	/organism="Salmonella enteritidis"
AUTHORS	/mol_type="genomic DNA"
TITLE	/strain="SE30"
JOURNAL	/db_xref="taxon:592"
REFERENCE	/note="amplification product"
AUTHORS	1. 230
TITLE	/gene="agfa"
JOURNAL	<1...>230
REFERENCE	/gene="agfa"
AUTHORS	/codon_start=1
TITLE	/transl_table=1
JOURNAL	/product="SEF17 fimbria"
REFERENCE	/protein_id="AAA96671.1"
AUTHORS	/db_xref="GI:1293678"
TITLE	/translation="GNHKGGSPPDSTLTYOYGSAANAALQSDARKSETTITGS GNGCADVGGGADNSITLITONGFPNNATITD"

ORIGIN	Query Match	Best Local Similarity	50.2%;	Score 229;	DB 1;	Length 230;
	Matches 229;	Conservative	99.6%;	Pred. No. 5.3e-55;	Mismatches 1;	Indels 0;
					Gaps 0;	
QY	88	GGAATATCATTAACGGCGGGCGGACATATGTTCCGGCCCGAGACTCAAGCTTATGAGCATTTATCAG				147
Db	1	GGTAATCATTAANGCGGGCGGCGCAATAGTTCGGGCCCGGACTCAAGCTTATGAGCATTTATCAG				60
QY	148	TACGGTTCGGCTTAACGAGCGTGGCGCTTGTCTGTGAAAGCCGATGCGCGTAAATCTGAAAACGACC				207
Db	61	TACGGTTCGGCTTAACGAGCGTGGCGCTTGTCTGTGAAAGCCGATGCGCGTAAATCTGAAAACGACC				120
QY	208	ATTACCCGAGAGCGGTTATGTATACGCGCGCGCATGTAGGCCGAGGGTGCAGATTAATAGTACT				267
Db	121	ATTACCCGAGAGCGGTTATGTATACGCGCGCGCATGTAGGCCGAGGGTGCAGATTAATAGTACT				180
QY	268	ATTGAACTGACTCAGAAATGTTTTCAGAAATTAATGCAACCAATCGACCGAGTG				317
Db	181	ATTGAACTGACTCAGAAATGTTTTCAGAAATTAATGCAACCAATCGACCGAGTG				230

RESULT	12
LOCUS	ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
DEFINITION	E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
ACCESSION	X90754
VERSION	X90754.1 GI:1147558
KEYWORDS	csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene; orfC gene.
SOURCE	Escherichia coli
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
REFERENCE	1
AUTHORS	Hammart,M., Arngvist,A., Bian,Z., Olsen,A. and Normark,S. Expression of two csg operons is required for production of fibronectin- and congo red-binding curli polymers in Escherichia
JOURNAL	Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE	96414468
PUBMED	8817489
REFERENCE	2 (bases 1 to 4680) Hammart M, Direct Submission Submitted (11-AUG-1995) M. Hammart, Karolinska Institutet, Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm, SWEDEN
TITLE	
JOURNAL	
FEATURES	
source	location/Qualifiers 1..4680 /organism="Escherichia coli" /mol_type="genomic DNA" /strain="K12" /db_xref="taxon:562" /map="23.15 minutes" 1..3140 /organism="Escherichia coli" /mol_type="genomic DNA" /strain="WC4100" /db_xref="taxon:562" 3141..4680 /organism="Escherichia coli" /mol_type="genomic DNA" /strain="WJ10" /db_xref="taxon:562" <1..>4680 /injection_seq="IS2 (partial)" complement(133)..966 /gene="csgG" complement(133)..966 /gene="csgG" /codon_start=1
repeat_region	
gene	
CDS	

```

/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62277.1"
/db_xref="GI:1147559"
/db_xref="GOA:P52103"
/db_xref="SWISS-PROT:P52103"
/translation="MORLFLVAVMLSSGLTAPKREARPTLMPRAQSYKDLTHLPA
PYGKIFVSVNYIOETGPKFYPASNPSTAPQSATMLYALKDSRFITLEKQGLQ
NLMBRKIRRAOENGTAINNR1PLOSITANIMVBSITIGESNVKSGGVARFEG
IGADTQYQDQIAVLRVNVSTGELISSVNTSKTILSYEQAGVFRIIDYQRLLEG
VGYTSNEPVMCLMSALETGVIPLINDIDRGLMDLQNKARQNDILVKYHMSVPE
S"
complement(993..1409)
/gene="csgB"
complement(993..1409)
/gene="csgB"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62278.1"
/db_xref="GI:1147560"
/db_xref="SWISS-PROT:P52104"
/translation="MRVHAVVLLMLISPLSMAGMTPEFQNPNGNNGAFILNS
AQONSYPNDPNDPFIETPSALDNFOAQIOQLIGSLISNINMGKRGKRVNTDIYV
DIARDQQLQANLDRKTCGSTIQVSGLOKNSIDF"
complement(1434..1823)
/gene="csgE"
complement(1434..1823)
/gene="csgE"
/codon_start=1
/transl_table=1
/product="assembly /transport component in curli
production"
/protein_id="CA62279.1"
/db_xref="GI:1147561"
/db_xref="SWISS-PROT:P52105"
/translation="MKRYLRWIVAAEPIFAAGNHAHVEVPGLLTDHTYSSIGHDY
RAFDKRESDTGNLTINERPSARWGSHITTVQDV1PQTFPLKRDPEKTIVFAL
IQTEALNRQIQALSTGDLADEF"
complement(1828..2478)
/gene="csgD"
complement(1828..2478)
/gene="csgD"
/note="necessary for transcription of the csgBA operon"
/codon_start=1
/transl_table=1
/product="putative regulatory protein"
/protein_id="CA62280.1"
/db_xref="GI:1147562"
/db_xref="GOA:P52106"
/db_xref="SWISS-PROT:P52106"
/translation="MFNEVSHIGHITLLITKSLQATATLQHLKQSLATIGKLIHQ
RSLDISSGSIILDMEMADKLIHYMODTSLKNNN1KILLNTPEDPYRDIENME
HINGVFVSMEDQEVNGLOGVLRGECYFOKLASVYLITSGNVRVSTSAITLTHE
KEIANKRIRICHSNNEIARSLPISNTYKTHLYNLFKIAVNR1QAVSWANDILNR"
complement(2626)
/note="transcriptional start site for csgDEFG operon"
3140
/note="transcriptional start site for csgBA operon"
3233..3688
/gene="csgB"
3233..3688
/gene="csgB"
/codon_start=1
/transl_table=1
/product="nucleation component of curli monomers"
/protein_id="CA62281.1"
/db_xref="GI:1147563"
/db_xref="GOA:P39828"
/db_xref="SWISS-PROT:P39828"
/translation="MKNKLLFMMLTIIGARIGAAAGVLANSEYNPAVNLKSPFN

```

```

OAA11GQGTNNNAQLRQSGSKLLAVVAOESSNRKXIDPDGNLAYIDQASANDA
SISQAGANTAMITQKSSKNKANTQGTQKTAIVQROSMALRVQR"
3729..4184
/gene="csgA"
3729..4184
/gene="csgA"
/note="major subunit"
/codon_start=1
/transl_table=1
/product="curliin/CsgA protein"
/protein_id="CA62282.1"
/db_xref="GI:1147564"
/db_xref="GOA:P28307"
/db_xref="SWISS-PROT:P28307"
/translation="WKLKVAIAIAIYFSGSALAGVYPQGGGNNHGGGNGSPNSE
LNIYQGGNSALALQTDARNSDITTIQHGKGNADVGQSDSSIDLTQKFGNSAT
LDQWNGKNSMNTVYQFGGNGAAYVDQASNSVNVTVQKGNNTAHQY"
3729..3788
/gene="csgA"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/transl_table=1
/product="putative curli production protein"
/protein_id="CA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLILAAISQITENTTQOQGVNTIPEVTLTQSLCRVQIL
SLREGSSGQSGQTKQEKTLSPANQPIALTKLSINISPDDBKVIIVTVSDQSLHSQO
WPPSSERS"

```

ORIGIN

Query Match 49.8%; Score 227.2; DB 1; Length 4680;
 Best Local Similarity 68.6%; Pred. No. 2.3e-54;
 Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTGCCTGCT 60
Db 3729 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTGTGGCAGTGCCTGCA 3788
QY 61 GCGCGTTCACAAATGGGCGGCGGCGGTATCATTAACGGCGGCAATAGTTCCGCG 120
Db 3789 GGTGTTTCTCTCACTACGCGCGGCGGCGGTACCAAGTGGCGGTAAATAGCGGC 3848
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACGCTGCGCTTCTCTGCA 180
Db 3849 CCAATTCGTAGCTGAACATTTACAGTACGTTGGCGGTAACTTGCACCTTCTCTGCA 3908
QY 181 AGCGATCCCGTAAATTTGAAAGCAGCATTAACCAAGCGGTTATGTTAAGCGCGCAT 240
Db 3909 ACTGATCCCGTAACTCTGACTGACTTAAACCAAGATGGCGGCTTAATGTTGACAGAT 3968
QY 241 GTAGGCGAGGGTGGGATTAATGACTTATGAAGTGAAGTGAAGTGGTTTCAGAAATAT 300
Db 3969 GTTGTAGAGCTGAGTATGACAGTCAATGATGACCAAGCTGCTTCTGTTAAGC 4028
QY 301 GCCACATGACAGTGAACGCTAAACCTCCGATTAATCTGTCGCGCAATATGATCAG 360
Db 4029 GCTACTTGTATCAGTGAACCGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGAT 4088
QY 361 CTGGTTACCCGTGTTTATCCCATGAATGACATGACAGCTTAATGTTGCGTCAAGTT 420
Db 4089 GGCACGCTGCTGAGTTGACAGACTGACATTAATCTCCCGTCAACGTGACTCAGGTT 4148
QY 421 GGTTTGGACACAGCGGCTAACAGATTTAA 456
Db 4149 GCGTTTGTAAACAGGACCGCTCATCACTACTAA 4184

```

RESULT 13
 AE000205

```

LOCUS       AE000205               10346 bp    DNA             linear    BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION   AE000205 U00096
VERSION     AE000205.1 GI:1787265
KEYWORDS
SOURCE
ORGANISM    Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
REFERENCE   1 (bases 1 to 10346)
AUTHORS     Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
            Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
            Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
            Mau, B. and Shao, Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5311), 1453-1474 (1997)
TITLE       JOURNAL
MOLTYPE     MEDLINE
PUBMED      97426617
            9278503
REFERENCE   2 (bases 1 to 10346)
AUTHORS     Blattner, F.R.
TITLE       Direct Submission
JOURNAL     Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3 (bases 1 to 10346)
AUTHORS     Blattner, F.R.
TITLE       Direct Submission
JOURNAL     Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecolli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4 (bases 1 to 10346)
AUTHORS     Plunkett, G. III.
TITLE       Direct Submission
JOURNAL     Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the database is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M55. No
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             location/Qualifiers
     source           1..10346
                     /organism="Escherichia coli K12"
                     /mol_type="genomic DNA"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
                     complement(1..48)
     repeat_region    complement(1..48)
     gene             1..5
                     /note="IS3"
                     /note="factor Sigma70; predicted +1 start at 1094689"
                     5..32
                     /note="factor Sigma70; predicted +1 start at 1094716"
                     69..392
                     /gene="b1028"
                     69..392
                     /gene="b1028"
                     /function="orf; Unknown"
                     /note="o107; This 107 aa ORF is 27 pct identical (0 gaps)
                     to 58 residues of an approx. 1868 aa protein FAS2_CANAL
                     SW: P43098"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74113.1"
                     /db_xref="GI:1787266"
                     /translation="MEDGVQAMRDYLAGDIDASPETHQVLMNVYAKSEVAPSIKENTLS
                     LHLTHVKMTBSLDITPLNMPVAFLEISNKPYLGNMINDFAGVDQQRVHRCRAFSQ
                     AKYFK"
                     275..299
                     /gene="b1028"
                     /note="factor Sigma32; predicted +1 start at 1094983"
                     389..1375
                     /gene="ycdY"
                     /note="synonym: b1029"
                     389..1375
                     /gene="ycdY"
                     /function="orf; Unknown"
                     /note="g328; This 328 aa ORF is 19 pct identical (4 gaps)
                     to 191 residues of an approx. 288 aa protein Y320_MYCB
                     SW: P47562"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74114.1"
                     /db_xref="GI:1787267"
                     /translation="MIPDYLPFRFODKRLIYIYVAILGLIGPYWKNAGPTPSSEDI
                     GVSGIALVLYNFIPLDKAYKAYCTKNDSPKRNKHIEFLTPDVAFLS
                     LIMSASMGVYOLPISLVAFLISLSPVITLRLMRTSYKQVALSAVAKVYK
                     SLTRYVILSVGISVNVLLITSPRNSDSVTGQMLTFRSTYKVALILGCVLAIMLF
                     PLRSKRVPLGRFLDIDLPFSSSEALSTFPKPLMLRLFLIVIEVMVILVSVL
                     ATIVEMKIVFEAFLLCYVPLIYFFYCRFLMNDPFMACDWFYFMGHFNK"
                     1253..1280
                     /gene="ycdY"
                     /note="central position to predicted promoter: -104"
                     1253..1280
                     /bound_moiety="Deor predicted site"
                     /gene="ycdY"
                     /note="No predicted promoter"
                     /bound_moiety="Deor predicted site"
                     1337..1365
                     /gene="ycdY"
                     /note="factor Sigma70; predicted +1 start at 1096049"
                     1494..1745
                     /gene="b1030"
                     1494..1745
                     /gene="b1030"
                     /function="orf; Unknown"
                     /note="o03; This 83 aa ORF is 25 pct identical (7 gaps)
                     to 81 residues of an approx. 616 aa protein ALBU_RABIT SW:
                     P49065"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC74115.1"
                     /db_xref="GI:1787268"
                     /translation="WGGRRCVGYTKKKVLSYELFFYGGEGGIDLSRSCGPAPAC
                     VHGLSNLSPVVDPRSGLLIPVCAIHKKKRTYRALL"
                     1513..1926
                     /gene="ycdY"
                     /note="synonym: b1031"

```

```

CDs
1513..1926
/gene="ycdV"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein R337_YEAST SM:
P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSPYFRTSSILMAVRGGLTRFAPAGSLPTAFVCP
GROLSTPGRSPRCVQYTKKPVLSLELFFKGGGGLDLSRSPCGAHCVHGL
SNMLSPVEPRSGLLIPVCAIYKKAFTFALL"
/complement(2111..2198)
/gene="serX"
/note="synonym: b1032"
/complement(2111..2198)
/gene="serX"
/product="tRNA-Ser"
/note="anticodon: CGA; CG Site No. 17878"
/function="tRNA; tRNA"
2269..2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338..2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355..2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393..3370
/gene="ycdW"
/note="synonym: b1033"
2393..3370
/gene="ycdW"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPRA_CORGL
SM: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="MRINFAORFSGVSMIDIFYPPTDTQWIEALRKAIPOARVAR
KSGNDSDADYALVWHPVEMLAGDIAKAPALGVDUSLTKQAHPEMNPSPRAR
LEDTGSEOMOEIVASQVLFHFRFPDRIDQNSHNPUPREDFPTGILAGVL
GSRVQSLQTPREPLRCWSPTRKSPGVQSPAGEELSAFLSQCRVAILNLPNPEV
GIINQQLLEKLPGDAVILNLAGVHVEDLLALDSGXVGMALDVFNNEPLPESP
LMQHPRTITPVAITRPAVAEYISRTIAQLEKSERVCGVDPRARGY"
3278..3308
/note="ycdW"
/note="factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396..3423
/promoter

Query Match 49.8%; Score 227.2; DB 1; Length 10346;
Best Local Similarity 68.6%; Pred. No. 2.5e-54;
Matches 313; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 1 ATGAAATCTTTAAAGTGCAGATTCGACAGCAATCGAGTTCTGCGAGTCTTGAGCT 60
Db 8993 ATGAAATCTTTAAAGTGCAGATTCGACAGCAATCGATTCCTCGAGAGCTCTGGCA 9052

Qy 61 GGCCTCGTTCCAGCATGGGGCGGGCGGCGGTATCATTAACGGCGCGCATATGTTCCGGC 120
Db 9053 GGTCTTCTCTCGATACGGGGCGGGCGGTAAACACGGTGTGGCGGTAAATATAGCGGC 9112

Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTAGCGTTCCGCTACGCTGGCGCTCTCTGCA 180
Db 9113 CCAAAATCTGAGCTGAACATTTACAGTACGTTGCGGTATCTTGCACTTGCTCTGCA 9172

Qy 181 AGCGATGCCCGTAATCTGAAACGACATTAACCGAGCGGTTATGGTAACGGCGCGCAT 240

```

```

Db 9173 ACTGATGCCCGTAATCTGACTGACTATTATACCGAGATGGCGGCGGTATGTGCAGAT 9232
Qy 241 GTAGCCAGGGTGGGATATATAGTACTTGAAGTCTGAAATGTTTCAAAATATAT 300
Db 9233 GTTGATGAGGCTCAGATGACGAGCTCAATCATCTGACCCAACTGGCTCGTTAACAGC 9292
Qy 301 GCCACATCGACGAGTGAAGCGTAAACACCCGATTTATATGTCGGCCATPATGATCAG 360
Db 9293 GCTACTCTGATCATGAGTAACGCGAANAATTTCTAAATAGACGTTAAACATTCGCTGCT 9352
Qy 361 CTGGTACCCGCTGTTGTTTACCATGAATAATGACATGACGACGTAATGTCGTCAGGTT 420
Db 9353 GGCACAGGTGTGAGTTGACACACGACGACCTAATCTCTCCGTCACGTCAGTACAGTT 9412
Qy 421 GGTTTGGCAACAGCCAGCGCTAACCGATTTAA 456
Db 9413 GGTCTTGTTAAACAGCGACGCTCATCAGTACTAA 9448

RESULT 14
D90741
LOCUS
DEFINITION
D90741 AB001340
ACCESSION
D90741.1 GI:1651509
VERSION
KEYWORDS
Complete and shotgun sequencing; csgG; csgF; csgE; csgD; csgB;
csgA; ycdB; cIs; nov; mdoG.
SOURCE
Escherichia coli K12
Organism
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
REFERENCE
1 Oshima,T., Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
Ikemoto,K., Inada,T., Itoh,T., Kajihara,M., Kanai,K., Kashimoto,K.,
Kimura,S., Kitagawa,M., Makino,K., Masuda,S., Miki,T.,
Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nishio,H.,
Nishio,Y., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K.,
Wada,C., Yamamoto,Y., Yano,M., and Horichi,T.
A 718-Kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map
JOURNAL
MEDLINE
PUBMED
8905232
TITLE
The systematic sequencing of the Escherichia coli genome in Japan
AUTHORS
Unpublished
3 (bases 1 to 15047)
Mori,H.
Direct Submission
Submitted (29-JUL-1996) Hirotsada Mori, NARA Institute of Science
and Technology, Res. & Edu. Center for Genetic Info., 8916-5
Takayama, Ikoma, Nara 630-01, Japan
(E-mail:hmori@rnc.aisf-nara.ac.jp, Tel:81-7437-2-5660,
Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A.,
Horichi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S.,
Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S.,
Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K.,
Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K.,
Nakamura,Y., Nishimoto,H., Nishio,Y., Oshima,T., Saito,N.,

```

Sampei, G., Seki, Y., Tagami, H., Takemoto, K., Wada, C.,
Yamamoto, Y. and Yano, M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirokazu Mori
Address: NARA Institute of Science and Technology,
Ikoma, 630-01, Japan
E-mail: hmori@nslc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http://bwt.aist-nara.ac.jp

FEATURES
Source

1..15047
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/db_xref="taxon:83333"
/map="23.7-24.0 min"
/clone="Kohara clone #231"
/note="Nucleotide position 1097101-1112147 from the
initiation site of Theta (0 min.). This clone is from
Kohara lambda miniset library."
<1..276

CDS

/note="unnamed protein product; ORF_ID:0230#5"
/codon_start=1
/transl_table=1
/protein_id="BAA35822.1"
/db_xref="GI:4062605"
/translation="DLFSSSENALSTFPAPKPLMLRLFILVIEVMITLVSVALIVE
WRWFAAYFLICVPLCIYFFYCRFLMNDPMACDMYFRMGHFNK"
1152..2090

CDS

/note="ORF_ID:0231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=1
/product="Hypothetical 33.0 kd protein in prob-proc
intergenic region."
/protein_id="BAA35823.1"
/db_xref="GI:4062606"
/translation="MDIIFHPPTDQWMEALRKAIPQARVARKSGNDSDVALV
WHPVEMLAGRDILKAVFALGAVDSILSKQAPEMINPSVPLFLEDQMGOMQOEY
ALVQVLFRRPDYRQONSSHMOPLPEYHREDFTIGIGAGVIGSKVAOSICTMRF
PLRCWSTRKSPGVOSFAGRELSAFLSQCRTLINLPTWPTVGIINQOLIEKIPD
GAYLINARGVHVEDDILALDSCKKGMVLVPRNBPPLPSPPLMQRHRTITPHV
AATTPRAVEYISRTTIAQLDEKGERVCGQVDRARGY"
2145..2882

CDS

/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=1
/protein_id="BAA35824.1"
/db_xref="GI:4062607"
/translation="MIFVDLHMTVASTAVSTLSDYIAQAKQKIKLPAITDGHGDM
EDAPHHMFTNMRIPWRVVDGILGIEANIKNVGIDICSGKMDSDLLIAGHE
VPAPHPDKATNTQAMATIASGNVHIISHGNPKYIDYKAAEAAKKQVLALEINS
SPHRSKSEEDNCREVAARVADAGWVALGSDSHTAFTNGEPFECIKILDVADPFPER
ILNVSPRLNLFESRGMAPIAEFADL"
2906..3460

CDS

/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=1
/product="Hypothetical protein H11543"
/protein_id="BAA35825.1"
/db_xref="GI:4062608"
/translation="MNEFSILCRVGLSVYROPQDPLVLPFTLLREKLANMPLQ
DELTRLOQSCMTQVSADYNALFTIGDECAVPPYSAMWEGATEAEEVAFISERGMPL
ADTPADHIGTLILASWMLDOSTEDSEALETLFSEYLLPMCGAFLGKVAEATTPFW
RTMAPLTRAIAIMWDELDEDSSE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

/codon_start=1
/transl_table=1
/protein_id="BAA35826.1"
/db_xref="GI:4062609"
/translation="MCSISFMAFAFAIMKGMNLLSIATTTGILSGIMGVAVSLGL
SMAGFLGCTAYFACPOGGKGLAISATLLSGVMVMVVIYSGALAPHEIGVYTG
IVAFLMCIQAKQLLSFVPDGTGACATPAGODMVLVPSIALGLIPYAMKNSGLM
LAARSAKTAREOEIKNKA"
complement (4117..4950)
/gene="cs9g"
complement (4117..4950)
/gene="cs9g"
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs9g precursor."
/protein_id="BAA35827.1"
/db_xref="GI:4062610"
/translation="MORLEFLVAVNMLSGCLTAPPKRAAPPTLMPRAQSYKDLTHLP
PGKTFVSYVNIODEGQKPYPAASFNAVPOSAATMLYTALKSRPFLPERGLO
NLINRKIRBAROEQTVAIINRRILOSITANINVEGSIIGESNVSGGVGARFG
ICADYQYLDQDIAVNLRYVNVSTGEILSVNTSKTILTEVQAGVFRFDYQRLLEG
VGTNSNEPVMCLMSAIEGTVEFLINDGIDRLMDLONKAERONDILVKYRMSVPE
S"
complement (4977..5393)
/gene="cs9g"
complement (4977..5393)
/gene="cs9g"
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs9g precursor."
/protein_id="BAA35828.1"
/db_xref="GI:4062611"
/translation="MRVHAAVVLMLISPLSWAGTMTFPNPNPFGNPNNGAFLINS
AQONSYKDPNDPFGIETPSALDNFTQALQSOILGGLSLNINTGKGRMVTNDYIV
DIANDQOLQNLVDRKTGRTSTIQVSLQNNSTDF"
complement (5418..5807)
/gene="cs9g"
complement (5418..5807)
/gene="cs9g"
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
cs9g precursor."
/protein_id="BAA35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEFLLFAAGNLHAAVEVPGLLTDHTVSSIGHDPY
RASFQKWSDETGNLTINERSARWGSMTITVWQDVIFFQFLPLRLRDEKTYVFL
IOTEBALNRDIONALSTGLDARDER"
complement (5812..6462)
/gene="cs9g"
complement (5812..6462)
/gene="cs9g"
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable cs9g operon transcriptional regulatory
protein."
/protein_id="BAA35830.1"
/db_xref="GI:4062613"
/translation="MENEASISHTGTLITKSSIQATALLQHKOSLATITGKLIHQ
RSLDISGSGSITLLDMEADKLIHYQDTLSKNNIKLILNTPEDYRIENMNP
HINQVYSMEDQERVNGLOGVARGECYFQKLASVLIHSGVRYNNTSSTALLTHRE
KEILNKIRIGASNNELARSLISENTYKTHLVNLFKIAKNTQAVSWANDMLRR"
7217..7672

/gene="csb"
7217..7672

Query Match 49.8%; Score 227.2; DB 1; Length 15047;

Best Local Similarity 68.6%; Pred. No. 2.6e-54; Mismatches 143; Indels 0; Gaps 0;

Matches 312; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Best Local Similarity 68.4%; Pred. No. 5.5e-54; Mismatches 144; Indels 0; Gaps 0;

Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

Query 1 ATGAACTTTTAAAGTGGAGCATTCCAGCAATCGATGTTCTGGAGAGCTCTGCT 60

Db 1 ATGAACTTTTAAAGTGGAGCATTCCAGCAATCGATGTTCTGGAGAGCTCTGCTGCA 60

Query 61 GCGCTGTTTCCAAATGG 120

Db 61 GGTGTTGTTCTCAAGTACGG 120

Query 121 CCGACATCAAGTGGAGCATTATCAAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

Db 121 CCAATTTCTGAGTGAACATTTACAGTACGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 180

Query 181 AGCGATGCCGTAATCTGAAACGACATTAACCAAGAGGGGTTATGTAACGGCCGAT 240

Db 181 ACTGATGCCGTAATCTGAAACGACATTAACCAAGAGGGGTTATGTAATGTCAGAT 240

Query 241 GTAGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300

Db 241 GTAGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300

Query 301 GCCACATCGACAGTGGAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGT 360

Db 301 GCTACTCTTGATCAAGTGAACGCAAAATTTCTGAATGACGTTAAACAGTTCCGTTG 360

Query 361 CTGGTTACCGGTGTTGTTACCATGAATGGAATGGAATGGAATGGAATGGAATGGAAT 420

Db 361 GCGAACGGTGTGCAAGTTGACCAAGTCAATCTAATCTCTCCGTCACAGTCAAGTT 420

Query 421 GGTGGGCAACACGCGCAGCTAACCGATTTAA 456

Db 421 GGTGGGCAACACGCGCAGCTAACCGATTTAA 456

Query 421 GGTGGGCAACACGCGCAGCTAACCGATTTAA 456

Db 421 GGTGGGCAACACGCGCAGCTAACCGATTTAA 456

Search completed: March 15, 2004, 22:50:05
Job time : 1965.17 secs

RESULT 15
AX814811 456 bp DNA linear PAT 05-DEC-2003

LOCUS AX814811
DEFINITION Sequence 15 from Patent WO03064446.
ACCESSION AX814811
VERSION AX814811.1 GI:39104001

KEYWORDS
ORGANISM Escherichia coli
SOURCE Escherichia coli
ORGANISM Escherichia coli

REFERENCE
1 Bioerck, L., Olsen, A., Wikstroem, M. and Herwald, H.
TITLE Peptides
JOURNAL Patent: WO 03064446-A 15 07-AUG-2003;
Hansa Medical Research Aktiebolag (SE)
Location/Qualifiers

FEATURES
1..456
/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/note="unassigned protein product"

CDS
1..456
/codon_start=1
/transl_table=1
/protein_id="CAB83541.1"
/db_xref="GI:39104002"
/translation="MKLLKVEATIAIVFSGSALAGVVPYGGGNGHGGGNGGNSGPNSE
LNTYVGSGNSALALOTDARNSDLTTOHGGNGADVGCGSDSDSIDLTORGFNSKAT
LDMNGNSKNTVAQFGGNGAAVDOTASNSVAVTQVGFNNATAHQY"

ORIGIN
Query Match 49.5%; Score 225.6; DB 6; Length 456;

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds

(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 1 atggaacttttaaaagtcgc.....ccacgcgtaccagctatctaa 456

Sequence: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Scoring table: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*\n2: geneseqn1990s:*\n3: geneseqn2000s:*\n4: geneseqn2001as:*\n5: geneseqn2001bs:*\n6: geneseqn2002s:*\n7: geneseqn2003as:*\n8: geneseqn2003bs:*\n9: geneseqn2003cs:*\n10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	3 AAC64623	Aac64623 Agfa::PT3
2	404.8	88.8	456	2 AAQ87467	Aa87467 Agfa sequ
3	404.8	88.8	456	2 AAT74142	Aat74142 Salmone11
4	404.8	88.8	456	3 AAC64617	Aac64617 Salmone11
5	382.4	83.9	456	3 AAC64628	Aac64628 Agfa::PT3
6	376	82.5	456	3 AAC64622	Aac64622 Agfa::PT3
7	361.6	79.3	456	3 AAC64626	Aac64626 Agfa::PT3
8	355.2	77.5	456	3 AAC64625	Aac64625 Agfa::PT3
9	353.6	77.5	456	3 AAC64629	Aac64629 Agfa::PT3
10	352	77.2	456	3 AAC64624	Aac64624 Agfa::PT3
11	350.4	76.8	456	3 AAC64630	Aac64630 Agfa::PT3
12	350.4	76.8	456	3 AAC64627	Aac64627 Agfa::PT3
13	348.8	76.5	456	3 AAC64631	Aac64631 Agfa::PT3
14	287.2	63.0	361	2 AAQ73066	Aa73066 Agfa sequ
15	287.2	63.0	361	2 AAT74141	Aat74141 Salmone11
16	225.6	49.5	456	3 AAC64619	Aac64619 Escherich
17	225.6	49.5	9 ACF36153	ACF36153	Ac36153 E. coli C
18	218	47.8	646	2 AAQ62647	Aa62647 Fibronect
19	156	34.2	369	2 AAQ62646	Aa62646 FNB curli
20	49	10.7	78	3 AAC64610	Aac64610 Agfa (SEF
21	48.4	10.6	78	3 AAC64609	Aac64609 Agfa (SEF
22	48	10.5	48	3 AAC64621	Aac64621 Leishmani
23	44.8	9.8	72	9 ACF36148	Ac36148 E. coli C

C	24	44.8	9.8	2000	7 ADA71938	Ada71938 Rice gene
C	25	43.2	9.5	48	3 AAC64616	Aac64616 S. enteri
C	26	41.6	9.1	78	3 AAC64605	Aac64605 Sefa (SEF
C	27	40.8	8.9	100	7 ACD68809	Ac68809 E. coli K
C	28	40.6	8.9	78	3 AAC64606	Aac64606 Sefa (SEF
C	29	37.6	8.2	456	3 AAC64618	Aac64618 Salmone11
C	30	37.4	8.2	456	3 AAC64620	Aac64620 Escherich
C	31	35.4	7.8	100	7 ACD68807	Ac68807 E. coli K
C	32	35.4	7.8	3411	5 AAS88526	Aa88526 DNA encod
C	33	35.4	7.8	3412	5 AAS89144	Aa89144 DNA encod
C	34	35	7.7	100	7 ACD68808	Ac68808 E. coli K
C	35	34.8	7.6	1083	5 AAS76745	Aa76745 DNA encod
C	36	34.4	7.5	2751	6 ABO70750	Ab70750 listeria
C	37	34.2	7.5	2000	7 ADA71938	Ada71938 Rice gene
C	38	34	7.5	1344	2 AAX09007	Aa09007 Hyperbena
C	39	34	7.5	1344	3 AAD00669	Aa00669 Erwinia a
C	40	34	7.5	1344	3 AAA14939	Aa14939 DNA encod
C	41	34	7.5	1344	6 AAL41133	Aa41133 Hyperbena
C	42	34	7.5	1344	6 AAD29125	Aa29125 Erwinia a
C	43	34	7.5	1344	6 AAD27016	Aa27016 E. amylov
C	44	34	7.5	1344	6 ABL51711	Ab51711 Erwinia a
C	45	33.4	7.3	78	9 ACF36151	Ac36151 E. coli C

ALIGNMENTS

RESULT 1
AAC64623
ID AAC64623 standard; DNA; 456 BP.
XX
AC AAC64623;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#2 DNA sequence SEQ ID NO:13.
XX
XX Salmone11; agfa; chromosomal gene replacement; fimbria; epitope;
KW vaccine; immune response; immunogen; ds.
XX
XX Salmone11 enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CAN00356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UYVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36347.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX PT that encodes foreign epitope or antigen, expresses recombinant Agfa
XX PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
XX assembly system of strains of Salmone11, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the

CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal
 CC comprising separating an amino acid polymer comprising a recombinant AgfA
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Bacteroides* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. The
 CC useful for the expression of recombinant AgfA protein which is useful
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match Best Local Similarity 100.0%; Score 456; DB 3; Length 456;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
 QY 61 GGGGTGCTTCCCAATGGGCGCGCGCGTATCATTAACGGCGCGCAATAGTCCGCG 60
 DB 61 GGGGTGCTTCCCAATGGGCGCGCGCGTATCATTAACGGCGCGCAATAGTCCGCG 60
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCTCTGCA 180
 QY 181 AGCGATGCCGTTAAATCGAAACGACATTAACGAGCGGTTATGTAACGGCGCGAT 240
 DB 181 AGCGATGCCGTTAAATCGAAACGACATTAACGAGCGGTTATGTAACGGCGCGAT 240
 QY 241 GTAGGCCAGGGTGGGTAATAGTACTATTGAATGAGTCAAGATGGTTTCAAGAAATAT 300
 DB 241 GTAGGCCAGGGTGGGTAATAGTACTATTGAATGAGTCAAGATGGTTTCAAGAAATAT 300
 QY 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCGATTAATCTGCGCCAAATATGATCAG 360
 DB 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCGATTAATCTGCGCCAAATATGATCAG 360
 QY 361 CTGGTTACCGGTTGTTTACCATGAATGACATGCAAGCGTAAATGGTGGTCAAGTT 420
 DB 361 CTGGTTACCGGTTGTTTACCATGAATGACATGCAAGCGTAAATGGTGGTCAAGTT 420
 QY 421 GGTGTTGGCAACAACGCGCAGGCTTAACGATTTAA 456
 DB 421 GGTGTTGGCAACAACGCGCAGGCTTAACGATTTAA 456

RESULT 2
 ID AA087467 standard; DNA; 456 BP.

XX AA087467;
 XX 25-MAR-2003 (revised)
 XX 26-JUN-1995 (first entry)
 DE AgfA sequence.
 XX
 XX Salmonella; AgfA; vaccine; genetic immunization; ds.
 OS Salmonella.
 XX

FH Key Location/Qualifiers
 FT CDS 1..454
 FT /*tag= a
 FT /note= "AgfA"
 PN W09425598-A2.
 PD 10-NOV-1994.
 PP 26-APR-1994; 94MO-1B000207.
 PR 26-APR-1993; 93US-00054452.
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 PI (KING/) KING J.
 PI Kay WW, Collinson SK, Clouthier SC, Doran JL,
 DR WPI; 1994-358275/44.
 DR P-PSDB; AAR74625.
 XX
 XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
 PT strains, vector constructs, or compans. contg. fimbrial type proteins.
 PS Disclosure; Fig 7B; 95pb; English.
 CC The DNA encodes the *Salmonella* AgfA protein. The DNA and isolated
 CC proteins are used in genetic immunization and vaccine compositions,
 CC respectively, to elicit an immune response to *Salmonella* in animals (e.g.
 CC food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX
 XX Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match Best Local Similarity 88.8%; Score 404.8; DB 2; Length 456;

Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGAGTCTGGCT 60
 QY 61 GGGGTGCTTCCCAATGGGCGCGCGGTAATCATTAACGGCGCGCAATAGTCCGCG 120
 DB 61 GGGGTGCTTCCCAATGGGCGCGCGGTAATCATTAACGGCGCGCAATAGTCCGCG 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGCTAACGCTGCTCTGCA 180
 QY 181 AGCGATGCCGTTAAATCGAAACGACATTAACGAGCGGTTATGTAACGGCGCGAT 240
 DB 181 AGCGATGCCGTTAAATCGAAACGACATTAACGAGCGGTTATGTAACGGCGCGAT 240
 QY 241 GTAGGCCAGGGTGGGTAATAGTACTATTGAATGAGTCAAGATGGTTTCAAGAAATAT 300
 DB 241 GTAGGCCAGGGTGGGTAATAGTACTATTGAATGAGTCAAGATGGTTTCAAGAAATAT 300
 QY 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCGATTAATCTGCGCCAAATATGATCAG 360
 DB 301 GCCACCATGACCAAGTGAACGCTTAAATCTCCGATTAATCTGCGCCAAATATGATCAG 360
 QY 361 CTGGTTACCGGTTGTTTACCATGAATGACATGCAAGCGTAAATGGTGGTCAAGTT 420
 DB 361 AATAAGCGCGGTTGTTTACCATGAATGACATGCAAGCGTAAATGGTGGTCAAGTT 420
 QY 421 GGTGTTGGCAACAACGCGCAGGCTTAACGATTTAA 456
 DB 421 GGTGTTGGCAACAACGCGCAGGCTTAACGATTTAA 456

RESULT 3
 AAT74142

ID AAT74142 standard; DNA; 456 BP.
XX AAT74142;
AC
XX
XX 25-MAR-2003 (revised)
DT 29-SEP-1997 (first entry)
XX
XX Salmomella enteritidis 27655-3b agfa gene.
DE
XX Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
XX Salmomella enteritidis.
OS
XX Key Location/Qualifiers
FH 1.456
FT /tag= a
FT /label= agfa_gene_fragment
FT /transl_except= (pos:367..369,aa:Pro)
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WM, Doran JL;
XX
XX MPI: 1997-309886/28.
DR P-PSDB; AAM23570.
XX
XX Isolated Salmomella gene agfa - used for diagnosis of Salmomella or
PT enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 19-112; 85pp; English.
XX
XX The present sequence represents an isolated agfa gene derived from
CC Salmomella enteritidis 27655-3b. The nucleic acid can be used to provide
CC diagnostic assays for Salmomella and/or enteropathogenic bacteria of the
CC family Enterobacteria. It can also be used to provide proteins and
CC antibodies which can be used for assays. The nucleic acid sequence can be
CC used to provide probes or primers which can specifically hybridise to
CC nucleic acid molecules from greater than 99% of Salmomella strains that
CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC correct PF field.)
XX
SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
Query Match 88.8%; Score 404.8; DB 2; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.2e-121;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGATGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGATGTTCTGGAGTCTTGCT 60
QY 61 GGCCTGTTCCACATGCGGCGCGCGGCGGATCATTAACGCGCGGCAATAGTTCCG 120
DB 61 GGCCTGTTCCACATGCGGCGCGCGGCGGATCATTAACGCGCGGCAATAGTTCCG 120
QY 121 CCGGACTCAACGTTGAGCATTTACATGCGTTCCGCTAACGCTGCGCTTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTACATGCGTTCCGCTAACGCTGCGCTTCTGCAA 180
QY 181 AGCGATCCCGTAAATCTGAAACGACCAATCCAGAGCGGTTATGTTAAAGCGCGCAT 240
DB 181 AGCGATCCCGTAAATCTGAAACGACCAATCCAGAGCGGTTATGTTAAAGCGCGCAT 240
QY 241 GTAGGCCAGGTCGGATTAATAGTACTATTGAACGACTCAGATGTTTCAAAATAT 300

DB 241 GTAGGCCAGGTCGGATTAATAGTACTATTGAACGACTCAGATGTTTCAAAATAT 300
QY 301 GCCACATGACGACGATGAAAGCTTAAATCTCGATATTACTGTGCGCAATATGATCAG 360
DB 301 GCCACATGACGACGATGAAAGCTTAAATCTCGATATTACTGTGCGCAATATGATCAG 360
QY 361 CTGGTACCCGCTGTTTACCATGAAATGACACATGCAACCGTAAATGCTGCTAGATT 420
DB 361 AATACCGCGCGCTGTTTAAATCAGACCGCATCTGATTCACCGCTAAATGTCGTCAGATT 420
QY 421 GATTGGCAACAACGACGCGCTAACAGTATTAA 456
DB 421 GATTGGCAACAACGACGCGCTAACAGTATTAA 456
RESULT 4
AAC64617
ID AAC64617 standard; DNA; 456 BP.
XX
XX AAC64617;
XX
XX 26-FEB-2001 (first entry)
XX
XX Salmomella enteritidis Agfa DNA sequence SEQ ID NO:1.
DE
XX Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
OS
XX WO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collinson SK, Kay WM;
XX
XX MPI: 2000-672631/65.
DR P-PSDB; AAB36341.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 134; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TRP) nucleation depended
CC assembly system of strains of Salmomella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CagA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmomella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fibribrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match	88.8%	Score 404.8	DB 3	Length 456
Best Local Similarity	93.0%	Pred. No. 1.2e-121		
Matches 424	Conservative	0	Mismatches 32	Indels 0
			Gaps	0

QY 1 ATGAACCTTTTAAAGTGGCAGCAATTCGACGATCGTAGTTTCGCGAGTGGCT 60

Nb 1 ATGAACCTTTTAAAGTGGCAGCAATTCGACGATCGTAGTTTCGCGAGTGGCT 60

61 GGCCTGTTCCACATGGGGCGGCGGTATCATTAACGCGCGGCAATAGTCCGAC 120

QY
121 CCGGACTCAACGTTCAGCATTTATCATGTAGCGTCCGGCTAACGCTGGCGCTTGCTTCGCAA 180

181 AGCGATGCCCGTAAATCTGAAAAGCAACCATTTACCCAGAGCGGTTATGGTAAAGGAGGGCGGAT 240

Db 181 AGCGATGCCCGTAAATCTGAAACGACCCATTACCCAGAGCGGTATGTGTAACGGCGCCGAT 240

241 GTAGGCAGGGGTCCGATATAGTACTATTGAACTGACTCAGAAATGGTTTCAGAAATAAT 300

DB
301 GCCACCATCGACCAGTGGAAACGTAACCTCGATATTACTGTGCGCCATATCGCGGT 360

361 CIGTATCCCGGTGTTATCCCATGAATGGCATGCAAGCGTATGTGCGCTAGGTT 420
361 AATATACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGACGCGTATGGTGGCTTAGGTT 420

421 GGTTTGGCAACAGCGCTAACGATATTA 456

RESULT 5

DD AAC64628 standard; DNA; 456 BP.
XX
AAC64628:

26-FEB-2001 (first entry)

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope; vaccine; immune response.

Salmonella enteritidis.
Escherichia coli.

WO200060103-A2.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.
(HVT-) INITI WTCORPTA

White AP, Doran JL, Collison SK, Kay WW;

XX
DR WPI: 2000-672631/65.
DR P-PsDE; AAB36352.
XX

recombinant agla gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.

XX The present invention describes a recombinant *agfa* gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a protein having a different function than

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;
SQ

Every match	83.9%;	Score 382.4;	DB 3;	Length 456;
Best Local Similarity	89.9%;	Pred. No. 2.4e-114;		
Matches 410; Conservative	0;	Mismatches 45;	Totals 0;	Cross 0;

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGAACTGATGTTCTGGCAGTGCCTGGCT 60

61 GCGCTCGTTCACACATGGGCGCGCGGTAATCATTAACGCGCGGCGCATAGTTCGGC 120

QY 121 CCGGACTCAACGTTGAGCAATTATCAGTACGGTCCGCTAACGCTGGCTGCTCGCAA 180

181 AGCGATGCCCCGTTAAATCTGAACGACCAATTACCCAGAGCGGTTATGTTAAACGGCGCCGAT 240
OY
-----CCGCGCCTTCTGTGCA 180

[illegible]

Db 241 GTAGCGCAGGGTGGGATATAGTACTATTGAAGTCACTGAGATGTTTCAGAAATAT 300

DB
301 GCCACCATCGACCGAGTGAACCTAAAAATATATACAGCTGTTACCCGTTGTTACC 360

Db

361 CATGAATGGACATGCAAAATAGAGCCGCATCTTATTTCCAGCGGTATGGTCCGTACAGTT 420

Db 421 GGTTCGCAACGACGCTAACCACTATTA 456

RESULT 6
AAC64622 standard; DNA; 456 BP.

XX ID AAC64622 standard; DNA; 456 BP.
XX
XX AAC64622;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#1 DNA sequence SEQ ID NO:11.
XX
XX Salmoneilla; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmoneilla enteritidis.
XX Escherichia coli.
XX Synthetic.
XX MO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36346.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 135; 139pp; English.

XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmoneilla, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CseA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmoneilla, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (I) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

XX
XX Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 82.5%; Score 376; DB 3; Length 456;
Best Local Similarity 89.0%; Pred. No. 2.9e-112;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGACGATTCGACGAATCGTAGTTTCGCGAGTCTCGCT 60
Db 1 ATGAACCTTTAAAGTGGACGATTCGACGAATCGTAGTTTCGCGAGTCTCGCT 60
QY 61 GCGGTGTTCCACATGAGGCGCGCGGTATCATCAAGCGCGGCAATAGTCCGAGC 120
Db 61 GCGGTGTTCCACATGAGGCGCGCGGTATCATCAAGCGCGGCAATAGTCCGAGC 120
QY 121 CCGGACTCAACGTTGAGCAATTCAGTACGCGTTCCGCTAACGCTGCGCTTCGCA 180
Db 121 CCGGACTCAACGTTGAGCAATTCAGTACGCGTTCCGCTAACGCTGCGCTTCGCA 180
QY 181 AGCGATGCCCTTAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
Db 181 AGCGATGCCCTTAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
QY 241 GTAGGCGAGGTTGCGGATTAATAGTACTTATGTAAGTGAAGTGTTCAGAAATAT 300
Db 241 GTAGGCGAGGTTGCGGATTAATAGTACTTATGTAAGTGAAGTGTTCAGAAATAT 300
QY 301 GCCACCATGACCAAGTGGAAAGCTTAAACTCGATATTAAGTCCCAATATGATCAG 360
Db 301 GCCACCATGACCAAGTGGAAAGCTTAAACTCGATATTAAGTCCCAATATGATCAG 360
QY 361 CTGGTACCGCGTGTGTTTACCATGAATGACATGCAATGACGTTATGTCAGATT 420
Db 361 CTGGTACCGCGTGTGTTTACCATGAATGACATGCAATGACGTTATGTCAGATT 420
QY 421 GGTTCGCAACGACGCGCTAATGATGATGATGATGATGATGATGATGATGATG 456
Db 421 GGTTCGCAACGACGCGCTAATGATGATGATGATGATGATGATGATGATGATG 456

RESULT 7
AAC64626 standard; DNA; 456 BP.

XX ID AAC64626 standard; DNA; 456 BP.
XX
XX AAC64626;
XX
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT3#5 DNA sequence SEQ ID NO:19.
XX
XX Salmoneilla; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
XX Salmoneilla enteritidis.
XX Escherichia coli.
XX Synthetic.
XX MO200060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000MO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UTVI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
XX P-PSDB; AAB36350.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.

XX
XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Agfa*, *CsgA* and *Agfa*-homologue fimbriae comprising recombinant
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant *Agfa*
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant *Agfa* protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match Best Local Similarity 79.3%; Score 361.6; DB 3; Length 456;
 Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTTGCGACATGTAAGTTCTGGAGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGCGACATTTGCGACATGTAAGTTCTGGAGTCTGGCT 60
 QY 61 GGGCGTTCGACCAATGGGCGGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 DB 61 GGGCGTTCGACCAATGGGCGGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCGATGCGTTCCGTAACGCTGCTGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCGATGCGTTCCGTAACGCTGCTGCTGCA 180
 QY 181 AGCGATCCCGTAATCTGAAACGACATTTACCCAGAGCGGTAATGTAACGCGCGAT 240
 DB 181 AGCGATCCCGTAATCTGAAACGACATTTACCCAGAGCGGTAATGTAACGCGCGAT 240
 QY 241 GTAGGCGAGGTCGCGATTAATGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 300
 DB 241 GTAGGCGAGGTCGCGATTAATGTAATCTGTAATCTGTAATCTGTAATCTGTAAT 300
 QY 301 GCCACCATCGACCGATGGAAGCTTAAACTCCGATTAATCTGCGCAATATGATCAG 360
 DB 301 GCCACCATCGACCGATGGAAGCTTAAACTCCGATTAATCTGCGCAATATGATCAG 360
 QY 361 CTGGTTCACCGTGTGTTTACCATGGAATGCGCATGAGTATGAGTGTGAGGTT 420
 DB 361 CTGGTTCACCGTGTGTTTACCATGGAATGCGCATGAGTATGAGTGTGAGGTT 420
 QY 421 GGTITGGCAACACGCGGCGGTAATCAAGCCATGATTCAGGTAATGAGTGTGAGGTT 480
 DB 421 GGTITGGCAACACGCGGCGGTAATCAAGCCATGATTCAGGTAATGAGTGTGAGGTT 480
 QY 481 GGTITGGCAACACGCGGCGGTAATCAAGCCATGATTCAGGTAATGAGTGTGAGGTT 540
 DB 481 GGTITGGCAACACGCGGCGGTAATCAAGCCATGATTCAGGTAATGAGTGTGAGGTT 540

RESULT 8
 AAC64625
 ID AAC64625 standard; DNA; 456 BP.
 AC AAC64625;
 XX
 DT 26-FEB-2001 (first entry)
 XX

DE *Agfa*: PT3#4 DNA sequence SEQ ID NO:17.
 XX *Salmonella*; *agfa*; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN MO20060102-A2.
 PD 12-OCT-2000.
 XX 05-APR-2000; 2000MO-CA000356.
 XX 05-APR-1999; 99US-0127888P.
 XX (UYVI-) UNIV VICTORIA.
 PI White AP, Doran JL, Collison SK, Kay W,
 DR WPI; 2000-672631/55.
 PS P-PSDB; AAB36349.
 PT Recombinant *agfa* gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant *Agfa*
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant *agfa* gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TA) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Agfa*, *CsgA* and *Agfa*-homologue fimbriae comprising recombinant
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant *Agfa*
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant *Agfa* protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 SQ Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;
 Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTTCCGACATTCGTAAGTTCTGGAGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGCGACATTTCCGACATTCGTAAGTTCTGGAGTCTGGCT 60
 QY 61 GGGCGTTCGACCAATGGGCGGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 DB 61 GGGCGTTCGACCAATGGGCGGCGGTAATCATTAAGGCGGGAATGTTCCGGC 120
 QY 121 CCGGACTCAACGTTGAGCATTTATCGATGCGTTCCGTAACGCTGCTGCTGCA 180
 DB 121 CCGGACTCAACGTTGAGCATTTATCGATGCGTTCCGTAACGCTGCTGCTGCA 180

```
Db 121 CCGGACTATGATCAAGCTGTTACCGCTGTTGTTACCCATGAATGGCAGCATGCTGCA 180
Oy 181 AGCGATGCGCGTAATCTGAACGACCATTCACGAGCGGTTATGTAAGCGCCCAT 240
Db 181 AGCGATGCGCGTAATCTGAACGACCATTCACGAGCGGTTATGTAAGCGCCCAT 240
Oy 241 GTAGGCGGAGGTCGGATATAGTACTATTGTAAGTCACTGAGAAATGTTTCAGAAATAT 300
Db 241 GTAGGCGGAGGTCGGATATAGTACTATTGTAAGTCACTGAGAAATGTTTCAGAAATAT 300
Oy 301 GCCACCATGACACGATGGAAGCTTAAATCTCCGATATTACTGCGCCAAATATGATCAG 360
Db 301 GCCACCATGACACGATGGAAGCTTAAATCTCCGATATTACTGCGCCAAATATGATCAG 360
Oy 361 CTGCTTACCGCTGTTGTTACCATGAATGGCAGCATGCAAGCGTAAATGTCGTCAGGTT 420
Db 361 AATTAACGCGCGCTGCTGTTAATCAAGCCGATCTGATTCACGCTAATGTCGTCAGGTT 420
Oy 421 GGTTCGCAACAAACGCGACGCTAACGATATTAA 456
Db 421 GGTTCGCAACAAACGCGACGCTAACGATATTAA 456

RESULT 9
AAC64629 standard; DNA; 456 BP.
XX
AC AAC64629;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#8 DNA sequence SEQ ID NO:25.
XX
KM Salmoneila; agfA, chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmoneila enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
FN WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000WO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WM;
XX
DR WPI; 2000-672631/65.
XX
DR P-PSDB; AAB36353.
XX
PT Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 138; 139p; English.
XX
XX The present invention describes a recombinant agfA gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SFB17/TAF) nucleation depended
XX assembly system of strains of Salmoneila, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX AgfA, CsgA and AgfA-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant AgfA
```

```
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmoneila, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
XX
SQ Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;
XX
Query Match 77.5%; Score 353.6; DB 3; Length 456;
XX Best Local Similarity 86.0%; Pred. No. 6.1e-105;
XX Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
Oy 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAGCAATCGTAGTTTCTGCGAGTCTTGCT 60
Oy 61 GCGCGTTCCTCCACATAGGGGGGGGGGGGTAATGATTAACGGCGGCGCAATAGTTCCGGC 120
Db 61 GCGCGTTCCTCCACATAGGGGGGGGGGGGTAATGATTAACGGCGGCGCAATAGTTCCGGC 120
Oy 121 CCGGACTCAACGTTGAGCATTATTCAGTACGGTTCCGCTAACGCTGCGCTTCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTATTCAGTACGGTTCCGCTAACGCTGCGCTTCTGCA 180
Oy 181 AGCGATCCCGTAATCTGAACGACCATTTACCGAGCGGTTATGTTACGCGCGCAT 240
Db 181 CTGCTTACCGCTGTTGTTTACCATGAATGGCAGATCGATGTTATGTTACGCGCGCAT 240
Oy 241 GTAGGCGGAGGTCGGATATAGTACTATTGTAAGTCACTGAGAAATGTTTCAGAAATAT 300
Db 241 GTAGGCGGAGGTCGGATATAGTACTATTGTAAGTCACTGAGAAATGTTTCAGAAATAT 300
Oy 301 GCCACCATGACACGATGGAAGCTTAAATCTCCGATATTACTGCGCCAAATATGATCAG 360
Db 301 GCCACCATGACACGATGGAAGCTTAAATCTCCGATATTACTGCGCCAAATATGATCAG 360
Oy 361 CTGCTTACCGCTGTTGTTACCATGAATGGCAGCATGCAAGCGTAAATGTCGTCAGGTT 420
Db 361 AATTAACGCGCGCTGCTGTTAATCAAGCCGATCTGATTCAGCGCTAATGTCGTCAGGTT 420
Oy 421 GGTTCGCAACAAACGCGACGCTAACGATATTAA 456
Db 421 GGTTCGCAACAAACGCGACGCTAACGATATTAA 456

RESULT 10
AAC64624 standard; DNA; 456 BP.
XX
AC AAC64624;
XX
DT 26-FEB-2001 (first entry)
XX
DE AgfA::PT3#3 DNA sequence SEQ ID NO:15.
XX
KM Salmoneila; agfA, chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmoneila enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
FN WO200060102-A2.
XX
```

PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI white AP, Doran JL, Collison SK, Kay WW;
 XX
 XI MPI; 2000-672631/65.
 XX
 DR P-PSDB; AAB36348.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 CC
 CC Sequence 456 BP, 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;
 XX
 SQ
 Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2e-104;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 1 ATGAACTTTTAAAGTGGAGCATTTGGACAAATCGTACGTTTCTGGCAGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTTGGACAAATCGTACGTTTCTGGCAGTCTGGCT 60
 QY 61 GGGGTCTTCCACAAATGGGCGGCGGGGATATCATTAACGGCGGCGCATTAAGTCCGGC 120
 DB 61 GGGGTCTTCCACAAATGGGCGGCGGGGATATCATTAACGGCGGCGCATTAAGTCCGGC 120
 DB 61 GGGGTCTTCCACAAATGGGCGGCGGGGATATCATTAACGGCGGCGCATTAAGTCCGGC 120
 QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACCGTGGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACCGTGGCTGCTCTGCA 180
 QY 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACCGTGGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTGACATTTATCAGTACGTTCCGCTAACCGTGGCTGCTCTGCA 180
 QY 181 AGCGATGCCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTGTAACGGCGCGAT 240
 DB 181 AGCGATGCCGTAATCTGAAACGACCATTAACCGAGCGGTTATGTGTAACGGCGCGAT 240
 QY 241 GTAGGCGGAGGTTGGGATTAATGACTGATGACTGATGACTGATGACTGATGACTGAT 300
 DB 241 GTAGGCGGAGGTTGGGATTAATGACTGATGACTGATGACTGATGACTGATGACTGAT 300
 QY 301 GCCACCATGACCAAGTGAACGCTAAACCTCGATATTAATGCTGGCCAAATGATCAG 360

DB 301 GCCACCATGACCAAGTGAACGCTAAACCTCGATATTAATGCTGGCCAAATGATCAG 360
 QY 361 GTGTTACCCGTTGTTTATCCATGAAATGACATGACGATTAATGTTGCTCAGGTT 420
 DB 361 AATAAGCCGCGCTGTTATCAGACCGCATCGATTCAGCGGTATGTTGCTCAGGTT 420
 QY 421 GGTTTGGCAACAGCGCGCTATACCGATTTAA 456
 DB 421 GGTTTGGCAACAGCGCGCTATACCGATTTAA 456
 RESULT 11
 AAC64630
 ID AAC64630 standard; DNA; 456 BP.
 XX
 AC AAC64630;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT39 DNA sequence SEQ ID NO:27.
 XX
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS *Salmonella enteritidis*.
 OS Synthetic.
 OS
 PN W0200060102-A2.
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI white AP, Doran JL, Collison SK, Kay WW;
 XX
 XI MPI; 2000-672631/65.
 XX
 DR P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

50 Sequence 456 BP, 115 A, 116 C, 118 G, 107 T, 0 U, 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GGGCGTCTTCCAAATGGGGGGCGCGGTATCTAATACGGCGGGAATAGTTCCGGC 120
DB 61 GGGCGTCTTCCAAATGGGGGGCGCGGTATCTAATACGGCGGGAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGATGCTTCCGCTAAAGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGATGCTTCCGCTAAAGCTGCTCTGCA 180
QY 181 AGCGATGCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAACGGCGCAT 240
DB 181 AGCGATGCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGACTATTGAACTGACTCAGATGTTTCAAAATAT 300
DB 241 TATATCATGCTGTGTACCGGCTGTGTATCCAGTAAATGGAGCATTCGTAATATAT 300
QY 301 GCCACATCGACAGTGGAAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
DB 301 GCCACATCGACAGTGGAAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
QY 361 CTGCTTACCGGTGTGTATTAATCCATGAAATGCAATGCAAGCGTATGCTGAGTT 420
DB 361 AATAACCGCGCGCTGTATTAATCAGACCGCATTCGATTCAGATGATGCTGAGTT 420
QY 421 GGTTTGGCAACAACGCGCATTAACAGATTTAA 456
DB 421 GGTTTGGCAACAACGCGCATTAACAGATTTAA 456

```

RESULT 12

AAC64627 standard; DNA; 456 BP.

```

XX AAC64627;
AC AAC64627;
XX 26-FEB-2001 (first entry)
DT XX
DE XX Agfa::PT3H6 DNA sequence SEQ ID NO:21.
XX XX
XX Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
OS Salmonella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX OS
XX MO20060102-A2.
XX PD
XX 12-OCT-2000.
XX PF 05-APR-2000; 2000WO-CA000356.
XX PR 05-APR-1999; 99US-0127888P.
XX PA (UVI-) UNIV VICTORIA.
XX PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI; 2000-672631/65.
DR

```

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP1/TFP) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, GsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

50 Sequence 456 BP, 112 A, 113 C, 125 G, 106 T, 0 U, 0 Other;

Query Match 76.8%; Score 350.4; DB 3; Length 456;
Best Local Similarity 85.5%; Pred. No. 6.8e-104;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTTCCAGCAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GGGCGTCTTCCAAATGGGGGGCGCGGTATCTAATACGGCGGGAATAGTTCCGGC 120
DB 61 GGGCGTCTTCCAAATGGGGGGCGCGGTATCTAATACGGCGGGAATAGTTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGATGCTTCCGCTAAAGCTGCTCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGATGCTTCCGCTAAAGCTGCTCTGCA 180
QY 181 AGCGATGCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAACGGCGCAT 240
DB 181 AGCGATGCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAACGGCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGACTATTGAACTGACTCAAGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGACTATTGAACTGACTCAAGATGTTTCAAAATAT 300
QY 301 GCCACATCGACAGTGGAAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
DB 301 GCCACATCGACAGTGGAAAGCTTAAATCCGATTTATGTCGCGCAATATGATCAG 360
QY 361 CTGCTTACCGGTGTGTATTAATCCATGAAATGCAATGCAAGCGTATGCTGAGTT 420
DB 361 AATAACCGCGCGTGTATTAATCAGACCGCATTCGATTCAGCGTAATGCTGAGTT 420
QY 421 GGTTTGGCAACAACGCGCATTAACAGATTTAA 456
DB 421 GGTTTGGCAACAACGCGCATTAACAGATTTAA 456

```

RESULT 13
AAC64631
ID AAC64631 standard; DNA; 456 BP.
XX
AC AAC64631;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#10 DNA sequence SEQ ID NO:29.
XX
KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX Escherichia coli.
XX Synthetic.
XX WO20060102-A2.
XX
XX 12-OCT-2000.
XX
XX 05-APR-2000; 2000WO-CA000356.
XX
XX 05-APR-1999; 99US-0127888P.
XX
XX (UUYI-) UNIV VICTORIA.
XX
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36355.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 138; 139pp; English.
XX
XX The present invention describes a recombinant agfa gene (1) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SFF17/TAFF) nucleation depended
XX assembly system of strains of Salmonella, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
XX protein containing a replacement segment or segments of foreign amino
XX acid sequence or sequences grown on a Salmonella, E. coli or
XX Enterobacteriaceae host cell, from the host cell and introducing the
XX polymer into the animal in conjunction with a carrier or diluent. (1) is
XX useful for the expression of recombinant Agfa protein which is useful for
XX eliciting an immune response in an animal. In a fimbrial presentation
XX system the heterologous antigens are presented in high numbers (up to
XX 500,000 copies/cell), the hybrid fimbria protein possesses both the
XX immunogenicity and adhesion properties relevant for an efficient live
XX vaccine, the carrier fimbrial subunit proteins are usually strong
XX immunogens, which may be important for directing an immune response
XX against the inserted epitope, and hybrid fimbriae are easy and
XX inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

SO Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.3e-103;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGACATCGTAGTTCTTGGCAGTGGCTGGCT 60

DB |||||
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGACATCGTAGTTCTTGGCAGTGGCTGGCT 60
QY 61 GGGCGTCCCAAAAGGGGGGGGGGGGATCATTAACGGCGGGCAATAGTTCGGC 120
DB 61 GGGCGTCCCAAAAGGGGGGGGGGGGATCATTAACGGCGGGCAATAGTTCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTCAACGGTGGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTTATCAGTACGGTTCGGTCAACGGTGGCTGCA 180
QY 181 AGCGATCCCGTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATCCCGTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCAGGAGTCCGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCAGGAGTCCGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACCGACGTAAGCGTAAACCTCCGATATTACTGCGCCATATGATCAG 360
DB 301 GCCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 CTGGTTACCCGCTGTTGTTACCATGAAATGCAATGCAAGGCTATGCTGATGATT 420
DB 361 AATPACGCCGCTGTTATATCAGACCGCATCTGATTCAGGCTAATGCTGCTGAGTT 420
QY 421 GGTGTTGGCAACAGCCGAGGCTAACCAGTATTA 456
DB 421 GGTGTTGGCAACAGCCGAGGCTAACCAGTATTA 456

RESULT 14
AAQ73066
ID AAQ73066 standard; DNA; 361 BP.
XX
XX AAQ73066;
XX
XX 27-AUG-2003 (revised)
XX DT 25-MAR-2003 (revised)
XX DT 26-JUN-1995 (first entry)
XX
XX Agfa sequence.
XX
XX Salmonella; agfa; vaccine; genetic immunization; ds.
XX
XX Salmonella enteritidis.
XX
XX Key Location/Qualifiers
XX CDS 1..359
XX /*tag= a
XX /*note= "Agfa"
XX FT /*tag= d
XX FT /*note= "TAF5 primer (pair with TAF6)"
XX FT 52..69
XX FT /*tag= b
XX FT /*note= "TAF3 primer (pair with TAF4)"
XX FT complement(103..129)
XX FT /*tag= e
XX FT /*note= "TAF6 primer (pair with TAF5)"
XX FT complement(129..402)
XX FT /*tag= c
XX FT /*note= "TAF4 primer (pair with TAF3)"
XX
XX WO9425598-A2.
XX
XX 10-NOV-1994.
XX PD 26-APR-1994;
XX PF 94WO-IB000207.
XX PR 26-APR-1993; 93US-00054452.
XX

PA	(UVV1-) UNIV VICTORIA INNOVATION & DEV CORP.
PA	(KING/) KING J.
PI	Kay WW, Collinson SK, Clouthier SC, Doran JL;
PI	WPI: 1994-358275/44.
DR	P-PsDB; AAR62761.
XX	Eliciting an immune response to Salmonella - using attenuated Salmonella
PT	strains, vector constructs, or compans. contg. fimbrial type proteins.
XX	Disclosure; Fig 7A; 95pp: English.
XX	The DNA encodes the Salmonella enteritidis27655-3b TnpH α mutant strain
CC	agfa gene cloned into pUC19. The DNA and isolated proteins are used in
CC	genetic immunization and vaccine compositions, respectively, to elicit an
CC	immune response to Salmonella in animals (e.g. food producing animals)
CC	and humane. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC	AUG-2003 to correct OS field.)
XX	Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
SO	
Query Match	63.0%; Score 287.2; DB 2; Length 361;
Best Local Similarity	99.0%; Pred. No. 2.7e-83;
Matches 289; Conservative	0; Mismatches 3; Indels 0; Gaps 0
QY	64 GTCTTCACATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGACCG 123
DB	1 GTCTTACACAGTGGGGCGCGCGGTATCATTAACGGCGCGCAATAGTCCGACCG 60
QY	124 GACTCAAGTTGAGCATTTATCAGTACGAGTTCGCTAACGCTGCCCTTCTCGCAAGC 183
DB	61 GACTCAAGTTGAGCATTTATCAGTACGAGTTCGCTAACGCTGCCCTTCTCGCAAGC 120
QY	184 GATGCCGCTTAATCTGAAACGACCTTACCCGACGCGTTATGTTAAGCGCGCATGTA 243
DB	121 GATGCCGCTTAATCTGAAACGACCTTACCCGACGCGTTATGTTAAGCGCGCATGTA 180
QY	244 GCGCAGGTTGCGGATATAGTACTTGTGACTGACTCAGAAATGTTTCAGAAATATGCC 303
DB	181 GCGCAGGTTGCGGATATAGTACTTGTGACTGACTCAGAAATGTTTCAGAAATATGCC 240
QY	304 ACCATGACCACTGGGAGCGCTTAACCTCCGATTTACTGTGCGCAATATG 355
DB	241 ACCATGACCACTGGGAGCGCTTAACCTCCGATTTACTGTGCGCAATATG 292
RESULT 15	
AAAT74141	
ID	AAAT74141 standard; DNA; 361 BP.
XX	AAAT74141;
AC	
XX	
DT	25-MAR-2003 (revised)
DT	30-SEP-1997 (first entry)
XX	
XX	Salmonella enteritidis 27655-3b TnpH α mutant agfa gene fragment.
DE	
XX	Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX	Salmonella enteritidis.
OS	
XX	
FH	Key Location/Qualifiers
FT	CDS 1..360
FT	/*tag= a
FT	/label= agfa_gene_fragment
FT	16..60
FT	/*tag= b
FT	/label= Primer_TAF5
FT	52..69
FT	/*tag= c
FT	/label= Primer_TAF3
FT	Complement(103..128)
FT	primer_bind

```

FT      /*tag= d
FT      /label= Primer_TAF6
FT      primer_bind
FT      complement(294..312)
FT      /tag= e
FT      /label= Primer_TAF4
XX      US5635617-A.
XX      PD
XX      03-JUN-1997.
XX      PF
XX      26-APR-1994; 94US-00233788.
XX      PR
XX      26-APR-1993; 93US-00054452.
XX      (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX      PA
XX      Collinson SK, Kay WM, Doran JL;
XX      WPI; 1997-309886/28.
XX      P-PSDB; AAM23569.
XX      Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX      enteropathogenic bacteria of the Enterobacteria family.
XX      Claim 1; Col 107-110; 85pp; English.
XX      PS
XX      The present sequence represents an isolated agfa gene fragment derived
XX      from Salmonella enteritidis 27655-3b Typhoid mutant strain. The nucleic
XX      acid can be used to provide diagnostic assays for Salmonella and/or
XX      enteropathogenic bacteria of the family Enterobacteria. It can also be
XX      used to provide proteins and antibodies which can be used for assays. The
XX      nucleic acid sequence can be used to provide probes or primers which can
XX      specifically hybridise to nucleic acid molecules from greater than 99% of
XX      Salmonella strains that are pathogenic to warm-blooded animals relative
XX      to nucleic acid molecules from virtually all other microbial organisms.
XX      (Updated on 25-MAR-2003 to correct P field.)
XX      CC
XX      SQ
XX      Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
XX
XX      Query Match 63.0%; Score 287.2; DB 2; Length 361;
XX      Best Local Similarity 99.0%; Pred. No. 2.7e-83;
XX      Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0
XX
XX      Db 64 GTCGTTCCACATGGGGCGGGCGGCGTATCATTAACGGCGCGGCAATATGTCGGCCCG 123
XX      1 GTCGTACACAGTGGGGCGGGCGGCGGCTATCATTAACGGCGCGGCAATATGTCGGCCCG 60
XX
XX      Oy 124 GACTCAACGTTGAGCAATTATTCAGTAGCGTTCCGCTAACGCTGCGCTTGTCTGCAAGC 183
XX      61 GACTCAACGTTGAGCAATTATTCAGTAGCGTTCCGCTAACGCTGCGCTTGTCTGCAAGC 120
XX
XX      Oy 184 GATGCCGCTAATTCGAAACGACATTAACCGAGAGCGGTTATGTTAAGGGCGCGATGTA 243
XX      121 GATGCCGCTAATTCGAAACGACATTAACCGAGAGCGGTTATGTTAAGGGCGCGATGTA 180
XX
XX      Oy 244 GGCCAGGGTGGCGAATATAGTACTATTGAATCACTGAGATGGTTTCAGAAATATATCC 303
XX      181 GGCCAGGGTGGCGAATATAGTACTATTGAATCACTGAGATGGTTTCAGAAATATATCC 240
XX
XX      Oy 304 ACCATCGACACAGTGAACGCTAAACCTCGATATTAAGTGTGGCCCAATATG 355
XX      241 ACCATCGACACAGTGAACGCTAAACCTCGATATTAAGTGTGGCCCAATATG 292
XX
XX      Search completed: March 15, 2004, 17:51:57
XX      Job time : 245.684 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgsc.....ccacgcgtacacgtattaa 456

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: Issued Patents NA: *
2: /cgn2_6/ptodata/2/ina/5A.COMB.seq: *
3: /cgn2_6/ptodata/2/ina/5B.COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6A.COMB.seq: *
5: /cgn2_6/ptodata/2/ina/6B.COMB.seq: *
6: /cgn2_6/ptodata/2/ina/BACKFILE1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	404.8	88.8	456	1	US-08-233-788A-58 Sequence 58, Appl
2	287.2	63.0	361	1	US-08-233-788A-56 Sequence 56, Appl
3	34	7.5	1344	3	US-09-120-927-1 Sequence 1, Appl
4	34	7.5	1344	4	US-09-431-614-5 Sequence 5, Appl
5	33	7.2	7766	4	US-09-125-619-3 Sequence 3, Appl
6	31.2	6.8	1491	6	5486473-3 Patent No. 5486473
7	31.2	6.8	1767	1	US-08-363-475-18 Sequence 18, Appl
8	31.2	6.8	2284	1	US-08-363-475-21 Sequence 21, Appl
9	31	6.8	1750	1	US-08-415-823-1 Sequence 1, Appl
10	31	6.8	1750	2	US-09-086-662-1 Sequence 15, Appl
11	30.6	6.7	1147	1	US-08-417-103-15 Sequence 5, Appl
12	30.6	6.7	1351	1	US-07-816-283-5 Sequence 5, Appl
13	30.6	6.7	1351	1	US-08-417-103-5 Sequence 5, Appl
14	30.6	6.7	1351	4	US-09-016-434-1303 Sequence 1303, Ap
15	30.6	6.7	1797	4	US-09-134-000C-1552 Sequence 1552, Ap
16	30.6	6.7	4403765	3	US-09-103-840A-2 Sequence 2, Appl
17	30.6	6.7	4411529	3	US-09-103-840A-1 Sequence 9, Appl
18	30.4	6.7	7183	4	US-09-081-149-9 Sequence 1, Appl
19	30.4	6.7	7183	4	US-09-081-149-10 Sequence 10, Appl
20	30.2	6.6	2007	4	US-09-643-597-153 Sequence 153, App
21	30.2	6.6	2007	4	US-09-480-884A-153 Sequence 153, App
22	30.2	6.6	2007	4	US-09-547-615A-153 Sequence 153, App
23	30.2	6.6	2007	4	US-09-606-421B-153 Sequence 153, App
24	30.2	6.6	2007	4	US-09-221-107-153 Sequence 153, App
25	30.2	6.6	2148	4	US-09-643-597-154 Sequence 154, App
26	30.2	6.6	2148	4	US-09-480-884A-154 Sequence 154, App
27	30.2	6.6	2148	4	US-09-547-615A-154 Sequence 154, App

C 28	30.2	6.6	2148	4	US-09-606-421B-154 Sequence 154, App
C 29	30.2	6.6	2148	4	US-09-221-107-154 Sequence 154, App
C 30	29.8	6.5	1008	4	US-09-252-991A-10904 Sequence 10904, A
C 31	29.8	6.5	1077	4	US-09-252-991A-10818 Sequence 10818, A
C 32	29.8	6.5	1521	4	US-09-252-991A-10865 Sequence 10865, A
C 33	29.8	6.5	3612	4	US-09-068-506-2 Sequence 2, Appl
C 34	29.8	6.5	3690	4	US-09-016-434-1234 Sequence 1234, Ap
C 35	29.8	6.5	4035	4	US-09-016-434-1369 Sequence 1369, Ap
C 36	29.6	6.5	972	4	US-09-328-352-2055 Sequence 2055, Ap
C 37	29.4	6.4	423	4	US-09-252-991A-7641 Sequence 7641, Ap
C 38	29.4	6.4	759	2	US-08-853-659A-18 Sequence 22, Appl
C 39	29.4	6.4	918	2	US-08-853-659A-22 Sequence 29, Appl
C 40	29.4	6.4	1252	4	US-09-561-756-29 Sequence 29, Appl
C 41	29.4	6.4	1252	4	US-09-227-721-29 Sequence 29, Appl
C 42	29.4	6.4	1252	4	US-09-954-697-29 Sequence 29, Appl
C 43	29.4	6.4	1939	4	US-08-961-527-310 Sequence 310, App
C 44	29.4	6.4	3660	4	US-09-252-991A-12269 Sequence 12269, A
C 45	29.4	6.4	9208	4	US-09-068-506-1 Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELECOMMUNICATION INFORMATION:
TELEFAX: (206) 682-6031
TELETYPE: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 88.8%; Score 404.8; DB 1; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.5e-11;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTAAAGTGCAGCATTCGCAAGTGTGTTCTGCGAGTCTGAGT 60
DB 1 ATGAACTTTTAAAGTGCAGCATTCGCAAGTGTGTTCTGCGAGTCTGAGT 60
QY 61 GGGGTGCTTCCACAAATGGGGCGGCGGATATCATACGGGGCGCAATAGTCCGGC 120
DB 61 GGGGTGCTTCCACAAATGGGGCGGCGGATATCATACGGGGCGCAATAGTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTTACGTAACGTTCCGTAACGTTGCGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTTACGTAACGTTCCGTAACGTTGCGCTCTGCA 180
QY 181 AGCGATGCCCGTAAATCTGAAAAGCAACCATTAACGAGGGGTTAGTAAAGCGCCGAT 240
DB 181 AGCGATGCCCGTAAATCTGAAAAGCAACCATTAACGAGGGGTTAGTAAAGCGCCGAT 240
QY 241 GTAGGCGAGGGTGGGATTAATAGTAACTGAACTGAGATGTTTCAAGAAATAT 300
DB 241 GTAGGCGAGGGTGGGATTAATAGTAACTGAACTGAGATGTTTCAAGAAATAT 300
QY 301 GGCACCATGACACGAGGAAAGCTTAAATCTCGATTTACTGTGCGCAATATGATCAG 360
DB 301 GGCACCATGACACGAGGAAAGCTTAAATCTCGATTTACTGTGCGCAATATGATCAG 360
QY 361 CTGATTACCCGTGTTGTTTACCATGAAATGCGACATGCAAGGTAATGTCGTCAGGT 420
DB 361 AATAAGCGCGGCTGTTTATATGAGACCGCATGTTCCAGCGTAAATGTCGTCAGGT 420
QY 421 GGTTTGGCAACAAAGCCGCGCTAACGATTTAA 456
DB 421 GGTTTGGCAACAAAGCCGCGCTAACGATTTAA 456

RESULT 2

US-08-233-788A-56

Sequence 56, Application US/08233788A

Patent No. 3633617

GENERAL INFORMATION:

APPLICANT: Dorian, James L.

APPLICANT: Kay, William W.

APPLICANT: Collinson, Karen S.

APPLICANT: Clouthier, Sharon C.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION

NUMBER OF SEQUENCES: 61

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,788A

FILING DATE: 26-Apr-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043, 403C2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-788A-56

Query Match 63.0%; Score 287.2; DB 1; Length 361;
Best Local Similarity 99.0%; Pred. No. 1,8e-90;
Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 64 GTGCTTCCACAAATGGGGCGGCGGATATCATACGGGGCGCAATAGTCCGGCCG 123
DB 1 GTGCTTCCACAAATGGGGCGGCGGATATCATACGGGGCGCAATAGTCCGGCCG 60
QY 124 GACTCAACGTTGAGCATTTTACGTAACGTTCCGTAACGTTGCGCTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCATTTTACGTAACGTTCCGTAACGTTGCGCTCTGCAAGC 120
QY 184 GATGCCGTTAAATCTGAAACGACATTAACGAGGGGTTAGTAAAGCGCCGATGTA 243
DB 121 GATGCCGTTAAATCTGAAACGACATTAACGAGGGGTTAGTAAAGCGCCGATGTA 180
QY 244 GGCAGGGTGGGATTAATAGTAACTGAACTGAGATGTTTCAAGAAATATGATCAG 303
DB 181 GGCAGGGTGGGATTAATAGTAACTGAACTGAGATGTTTCAAGAAATATGATCAG 240
QY 304 ACCATGACACGAGGAAAGCTTAAATCTCGATTTACTGTGCGCAATATGATCAG 355
DB 241 ACCATGACACGAGGAAAGCTTAAATCTCGATTTACTGTGCGCAATATGATCAG 292

RESULT 3

US-09-120-927-1/C

Sequence 1, Application US/09120927

Patent No. 6262018

GENERAL INFORMATION:

APPLICANT: Kim, Jihyun Francis

APPLICANT: Beer, Steven V.

TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

STREET: P.O. Box 1051, Clinton Square

CITY: Rochester

STATE: New York

COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/120,927

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/055,108

FILING DATE: 06-AUG-1977

ATTORNEY/AGENT INFORMATION:

NAME: Goldman, Michael L.

REGISTRATION NUMBER: 30,727

REFERENCE/DOCKET NUMBER: 19603/1581

TELECOMMUNICATION INFORMATION:

TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 1344 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match
Best Local Similarity 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGCTGCGCTGCTTCACAAATGGGCGCGCGGTAATCATPAACGGCGCGCAAT 111
Db 209 GCTCCGGTTCGCGCACTGATGTTGGCGATGACGATTAACAGTTCCGCCAAT 150
Qy 112 AGTTCGGCGCGCGCACTGACGTTGAGCATTTATGAGTTCGCGCTAAGCGCTT 171
Db 149 AATTGAGCCATTGCTCAATGGTTGGCGATCGATGGGTTGGCCCAAGCAAAATTT 90
Qy 172 GC 173
Db 89 GC 88

RESULT 4

US-09-431-614-5/c
Sequence 5, Application US/09431614
Patent No. 6624139
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Schading, Richard L.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
FILE REFERENCE: 21829/41 (EBC-003)
CURRENT APPLICATION NUMBER: US/09/431,614
CURRENT FILING DATE: 1999-11-02
EARLIER APPLICATION NUMBER: 60/107,243
EARLIER FILING DATE: 1998-11-05
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match
Best Local Similarity 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.13;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Qy 52 GCTCTGCTGCGCTGCTTCACAAATGGGCGCGCGGTAATCATPAACGGCGCGCAAT 111
Db 209 GCTCCGGTTCGCGCACTGATGTTGGCGATGACGATTAACAGTTCCGCCAAT 150
Qy 112 AGTTCGGCGCGCGCACTGACGTTGAGCATTTATGAGTTCGCGCTAAGCGCTT 171
Db 149 AATTGAGCCATTGCTCAATGGTTGGCGATCGATGGGTTGGCCCAAGCAAAATTT 90
Qy 172 GC 173
Db 89 GC 88

RESULT 5

US-09-125-619-3
Sequence 3, Application US/09125619
Patent No. 6437116
GENERAL INFORMATION:
APPLICANT: NORRIS, STEVEN J.
APPLICANT: JING-BEN, ZHANG
APPLICANT: HARDHAM, JOHN M.
APPLICANT: HOWELL, JERRILYN K.
APPLICANT: BARBOUR, ALAN G.
APPLICANT: WEINSTOCK, GEORGE M.

TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
FILE REFERENCE: UMSH-234
CURRENT APPLICATION NUMBER: US/09/125,619
CURRENT FILING DATE: 1999-01-27
NUMBER OF SEQ ID NOS: 48
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 3
LENGTH: 7766
TYPE: DNA
ORGANISM: Borrelia burgdorferi
FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-09-125-619-3

Query Match
Best Local Similarity 7.2%; Score 33; DB 4; Length 7766;
Best Local Similarity 47.8%; Pred. No. 0.77;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGGCTTCTCTGCAAGAGATGCCGTAATGTGAAGCAATTAACCG 216
Db 4432 GTTACGGCTGGCGCTGCTGTGACAGAGATGAGAGAGAGCGCTGACAGGCTTAACAAATCCG 4491
Qy 217 AGCGATTATGATAGCGCGCGGATGATGAGCGAGGATGCGATTAATGACTATGAACTG 276
Db 4492 ATTCTGCTGCTATTGGAGAGGATGATGAGATGATGAGATGATGATGATGATGATG 4551
Qy 277 ACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 336
Db 4552 AAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4511
Qy 337 ATTACTGTGCGCAATATGAT 357
Db 4612 TTTCCTGGAAGATGATGAT 4632

RESULT 6

5486473-3/c
Patent No. 5486473
APPLICANT: FUKITA, HIROYUKI; YOSHIDA, IWANO; TAKAGI, MITSUO;
MANABE, SADAOKI; FUKAI, KOSUKE
TITLE OF INVENTION: A DNA CODING FOR A FLAVIVIRUS ANTIGEN
NUMBER OF SEQUENCES: 16
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/194,049
FILING DATE: 09-FEB-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 809,255
FILING DATE: 18-DEC-1991
APPLICATION NUMBER: 279,685
FILING DATE: 05-DEC-1988
APPLICATION NUMBER: 932,419
FILING DATE: 19-NOV-1986
SEQ ID NO: 3
LENGTH: 1491
5486473-3

Query Match
Best Local Similarity 6.8%; Score 31.2; DB 6; Length 1491;
Best Local Similarity 50.0%; Pred. No. 1.3;
Matches 78; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 178 CAAGCGATGCCGCTTAATCTGAAGACCACTTAACCAAGCGGTTATGTAAGCGGCC 237
Db 204 CAAGCTACTGACCAAGCTAGTGAACGTAACGATGCGGACACATCTGCGAGGTGGCTGCTTC 145
Qy 238 GATGTAGCGCAGGATGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 297
Db 144 CATGTTATCATTTTGAACATCATGATGATGATGATGATGATGATGATGATGATGATGATG 85
Qy 298 AATGCAACATGACCAAGTGAAGCGTAAACATCC 333
Db 84 ATGCGCTTCCAGTACGATCAACCCATGTAGCTCC 49

RESULT 7

US-08-363-475-18/c

Sequence 18, Application US/08363475

Patent No. 5516679

GENERAL INFORMATION:

APPLICANT: Chiang, Shu-Jen

APPLICANT: Burnett Jr., William V.

APPLICANT: Tonzi, Sean M.

TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM

TITLE OF INVENTION: FUSARIUM OXYSPORUM

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thomas R. Savitsky

STREET: P.O. Box 4000

CITY: Princeton

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/363,475

FILING DATE: 23-DEC-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Savitsky, Thomas R.

REGISTRATION NUMBER: 31,661

REFERENCE/DOCKET NUMBER: ON-0134

TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 252 4956

TELEFAX: (609) 252 4526

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 1767 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..1521

US-08-363-475-18

Query Match
Best Local Similarity 55.6%; Score 31.2; DB 1; Length 1767;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 278 CTCGAAATGCTTTGAGAAATATGCGACATCGACAGTGAACGCTAAAACTCCGATA 337
DB 1391 CCCCCAATGCTTTCTGGCGCAACCTCTTCTCGACCCATTTTCACATFACCGAAGACACA 1332

QY 338 TTACTGTGGCGCAATATGATGATGCTGTACCCGCTGTGTGTTACCCATG 385
DB 1331 TTAGTATCTGCGCTCTTCCGCCCATGTTCCGGCGTTGTTCCATG 1284

RESULT 8

US-08-363-475-21/c

Sequence 21, Application US/08363475

Patent No. 5516679

GENERAL INFORMATION:

APPLICANT: Chiang, Shu-Jen

APPLICANT: Burnett Jr., William V.

APPLICANT: Tonzi, Sean M.

TITLE OF INVENTION: PENICILLIN V AMIDOHYDROLASE GENE FROM

TITLE OF INVENTION: FUSARIUM OXYSPORUM

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

ADDRESSEE: Thomas R. Savitsky

STREET: P.O. Box 4000

CITY: Princeton

STATE: New Jersey

COUNTRY: U.S.A.

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/415,823

ADDRESSEE: Thomas R. Savitsky
STREET: P.O. Box 4000
CITY: Princeton
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/363,475

FILING DATE: 23-DEC-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Savitsky, Thomas R.

REGISTRATION NUMBER: 31,661

REFERENCE/DOCKET NUMBER: ON-0134

TELECOMMUNICATION INFORMATION:

TELEPHONE: (609) 252 4956

TELEFAX: (609) 252 4526

INFORMATION FOR SEQ ID NO: 21:

SEQUENCE CHARACTERISTICS:

LENGTH: 2284 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURE:

NAME/KEY: CDS

LOCATION: 241..1836

US-08-363-475-21

Query Match
Best Local Similarity 55.6%; Score 31.2; DB 1; Length 2284;
Matches 60; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 278 CTCGAAATGCTTTGAGAAATATGCGACATCGACAGTGAACGCTAAAACTCCGATA 337
DB 1706 CCCCCAATGCTTTCTGGCGCAACCTCTTCTCGACCCATTTTCACATFACCGAAGACACA 1647

QY 338 TTACTGTGGCGCAATATGATGATGCTGTACCCGCTGTGTGTTACCCATG 385
DB 1646 TTAGTATCTGCGCTCTTCCGCCCATGTTCCGGCGTTGTTCCATG 1599

RESULT 9

US-08-415-823-1

Sequence 1, Application US/08415823

Patent No. 5759538

GENERAL INFORMATION:

APPLICANT: Donovan, William P.

APPLICANT: Tan, Yuying

TITLE OF INVENTION: BACILLUS THURINGIENSIS APT AND NPT

TITLE OF INVENTION: PROTEASE GENES AND ALKALINE PROTEASE DEFICIENT AND NEUTRAL

NUMBER OF SEQUENCES: 9

CORRESPONDENCE ADDRESS:

ADDRESSEE: Panitch Schwarze Jacobs & Nadel, P.C. c/o

STREET: 1601 Market Street, 36th floor

CITY: Philadelphia

STATE: Pennsylvania

COUNTRY: USA

ZIP: 19103-2398

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/415,823

Query Match	6.8%	Score 31;	DB 2;	Length 1750;
-------------	------	-----------	-------	--------------

Db 302 CAGACCCAGCATGAAAGAGTCTTCGCGATGCGCAGGTTGAGAGTGAATGTTGTGAT 243
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCGACAGTGAACGCTAAATACTCC 333
Db 242 GGTCTTCATCTTGCGATAGCGAGATGACATTAATGACAGTGTGTGCCACACACCC 183
QY 334 GATATTACTGTGCGGCATATGATCAGCTGTACCCGTGTGTTTACCATGAATGGCA 393
Db 182 AATGATGACAGACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGATGATGA 123
QY 394 C 394
Db 122 C 122

RESULT 12

US-07-816-283-5/C
; Sequence 5, Application US/07816283
; Patent No. 5436155
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Saino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: PO Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/816,283
; FILING DATE: 19911231
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: McDaniel, C. Steven
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-787-1400
; TELEFAX: 713-789-2679
; TELETYPE: 79-0924
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-07-816-283-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 214 CAGACCGGTATGTAACGGCGCGATGTAGGCGCAGGTCGCGATTAATGTAATTTGA 273
Db 370 CAGACCCAGCATGAAAGAGTCTTCGCGATGCGCAGGTTGAGAGTGAATGTTGTGAT 311
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCGACAGTGAACGCTAAATACTCC 333
Db 310 GGTCTTCATCTTGCGATAGCGAGATGACATTAATGACAACTGTGTGCCACACACCC 251
QY 334 GATATTACTGTGCGGCATATGATCAGCTGTACCCGTGTGTTTACCATGAATGGCA 393
Db 250 AATGATGACAGACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGATGATGA 191
QY 394 C 394

Db 190 C 190

RESULT 13

US-08-417-103-5/C
; Sequence 5, Application US/08417103
; Patent No. 5723299
; GENERAL INFORMATION:
; APPLICANT: Bell, Graeme I.
; APPLICANT: Yamada, Yuichiro
; APPLICANT: Saino, Susumu
; TITLE OF INVENTION: SOMATOSTATIN RECEPTORS
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/417,103
; FILING DATE: 05-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/816,283
; FILING DATE: 01-DEC-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mark B.
; REGISTRATION NUMBER: 37,259
; REFERENCE/DOCKET NUMBER: ARCD:144
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 83..1189
US-08-417-103-5

Query Match 6.7%; Score 30.6; DB 1; Length 1351;
Best Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 214 CAGACCGGTATGTAACGGCGCGATGTAGGCGCAGGTCGCGATTAATGTAATTTGA 273
Db 370 CAGACCCAGCATGAAAGAGTCTTCGCGATGCGCAGGTTGAGAGTGAATGTTGTGAT 311
QY 274 CTGACTCAGATGTTTTCAGAAATATCCACATCGACAGTGAACGCTAAATACTCC 333
Db 310 GGTCTTCATCTTGCGATAGCGAGATGACATTAATGACAACTGTGTGCCACACACCC 251
QY 334 GATATTACTGTGCGGCATATGATCAGCTGTACCCGTGTGTTTACCATGAATGGCA 393
Db 250 AATGATGACAGACCAAAATAGATGAATGTGAGAGTGCATTTGCTGTGATGATGA 191
QY 394 C 394
Db 190 C 190

RESULT 14

```

US-09-016-434-1303/c
; Sequence 1303, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
; APPLICANT: Janice Au-Young
; APPLICANT: Jeffrey J. Seilhamer
; TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
; TITLE OF INVENTION: PATHWAY GENE EXPRESSION
; NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
; STREET: 3174 PORTER DRIVE
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/016,434
; FILING DATE: HERewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Zeller, Karen J.
; REGISTRATION NUMBER: 37,071
; REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 855-0555
; TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 1303:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1351 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEBANK
; CLONE: G307435
; US-09-016-434-1303

Query Match 6 7%; Score 30.6; DB 4; Length 1351;
Beet Local Similarity 48.1%; Pred. No. 2;
Matches 87; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

QY 214 CAGAGCGGTTATGTTACGCGCGCGATGTAGCGCAGGGGCGGATAATAGTACTATTGAA 273
Db 370 CAGACCCAGCATGAAAGACTCATCTGCGATGCGCGGTTGAGATGTAAATGTTGTGTAAT 311
QY 274 CTGACTCGAAATGCTTTGAGAAATATATCCACCATCGACAGGTGMAAGCTTAAAACTCC 333
Db 310 GGTCTTCACTTTGGCATAGCGGAGATGACATTAATGACAAGTGTTGTTGCCACACACC 251
QY 334 GATATTACTGTCCGCGCAATATGATAGACTGTTACCGGTGTTGTTATCCATGAATGGCA 393
Db 250 AATGATGACAGCACCAAAATATGATGATGTGAGGACTGATTCCTGTTCAGGTCAATAGTA 191
QY 394 C 394
Db 190 C 190

RESULT 15
US-09-134-000C-1552/c
Sequence 1552, Application US/09134000C
Patent No. 6617156
; GENERAL INFORMATION:

```

```

APPLICANT: Lym Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 032796-032
CURRENT APPLICATION NUMBER: US/09/134,000C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/055,778
PRIOR FILING DATE: 1997-08-15
NUMBER OF SEQ ID NOS: 6812
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1552
LENGTH: 1797
TYPE: DNA
ORGANISM: Enterococcus faecalis
US-09-134-000C-1552

Query Match      6.7%; Score 30.6; DB 4; Length 1797;
Best Local Similarity 65.2%; Pred. No. 2.3;
Matches 45; Conservative 0; Mismatches 24; Indels 0; Gaps 0

QY      361 CTGGTACCCGTGTTGTTACCAATGAATGCAATGCATGCACGCTAATGCTGCATAGCTT 420
      |||||
DB      721 CTGGTTTCTTCTTCGCTTCATGAAATTCACATATAAGCGCAATATTCATTAATTG 662

QY      421 GGTTTGGC 429
      |||||
DB      661 GTTGTTGGC 653

```

Search completed: March 16, 2004, 04:37:25
Job time : 53.6647 secs

THIS PAGE BLANK (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds

(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgcgttaaccagttatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	34.4	7.5	2751	US-10-398-221-3563	Sequence 3563, Ap
2	34.2	7.5	616	US-10-424-599-88945	Sequence 88945, A
3	34	7.5	1344	US-09-835-684-6	Sequence 6, Appli
4	34	7.5	1344	US-09-880-371-6	Sequence 6, Appli
5	34	7.5	1344	US-09-879-248-5	Sequence 5, Appli
6	34	7.5	1344	US-10-010-390-6	Sequence 5, Appli
7	34	7.5	1344	US-10-441-736-5	Sequence 5, Appli
8	34	7.5	1625	US-10-369-493-26980	Sequence 26980, A
9	33.2	7.3	1662	US-10-156-761-3483	Sequence 3483, Ap
10	33.2	7.3	9025608	US-10-156-761-1	Sequence 1, Appli
11	33	7.2	7766	US-10-222-566-3	Sequence 3, Appli
12	33	7.2	7766	US-10-222-162-3	Sequence 3, Appli
13	33	7.2	7766	US-10-143-024-3	Sequence 3, Appli
14	33	7.2	368004	US-09-949-654-3	Sequence 3, Appli
15	32.8	7.2	1776	US-10-282-122A-24385	Sequence 24385, A

C 16	32.6	7.1	596	14	US-10-184-644-310	Sequence 310, App
C 17	32.6	7.1	596	14	US-10-184-634-310	Sequence 310, App
C 18	32.6	7.1	596	14	US-10-063-685-100	Sequence 100, App
C 19	32.2	7.1	568	14	US-10-333-631-1	Sequence 1, Appli
C 20	32.2	7.1	708	12	US-10-282-122A-27953	Sequence 27953, A
C 21	32.2	7.1	972	14	US-10-333-631-4	Sequence 4, Appli
C 22	32	7.0	983	9	US-09-764-864-465	Sequence 465, App
C 23	31.8	7.0	750	14	US-10-184-644-104	Sequence 104, App
C 24	31.8	7.0	750	14	US-10-184-634-104	Sequence 104, App
C 25	31.8	7.0	978	15	US-10-369-493-24285	Sequence 24285, A
C 26	31.8	7.0	4589	10	US-09-764-891-8845	Sequence 8845, Ap
C 27	31.8	7.0	4590	10	US-09-764-891-8846	Sequence 8846, Ap
C 28	31.8	7.0	24081	14	US-10-132-134-13	Sequence 13, Appli
C 29	31.8	7.0	52101	14	US-10-132-134-1	Sequence 1, Appli
C 30	31.6	6.9	4830	12	US-10-282-122A-27869	Sequence 27869, A
C 31	31.6	6.9	6003	13	US-10-010-501-9	Sequence 9, Appli
C 32	31.4	6.9	269	9	US-09-923-876-5903	Sequence 5903, Ap
C 33	31.4	6.9	269	11	US-09-923-876-5903	Sequence 5903, Ap
C 34	31.4	6.9	1287	15	US-10-369-493-46950	Sequence 46950, A
C 35	31.4	6.9	2007	12	US-10-425-114-31950	Sequence 31950, A
C 36	31.4	6.9	2074	12	US-10-425-114-27886	Sequence 27886, A
C 37	31	6.8	2886	14	US-09-801-368-131	Sequence 131, App
C 38	30.8	6.8	421	10	US-09-918-995-33997	Sequence 33997, A
C 39	30.6	6.7	1110	11	US-09-826-509-570	Sequence 570, App
C 40	30.6	6.7	1110	14	US-10-225-567A-313	Sequence 313, App
C 41	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
C 42	30.6	6.7	1351	15	US-10-305-720-1303	Sequence 1303, Ap
C 43	30.6	6.7	1510	15	US-10-292-798-677	Sequence 677, App
C 44	30.6	6.7	1510	15	US-10-292-798-677	Sequence 677, App
C 45	30.6	6.7	1608	9	US-09-815-242-3923	Sequence 3923, Ap

ALIGNMENTS

RESULT 1
US-10-398-221-3563
Sequence 3563, Application US/10398221
Publication No. US20040018514A1
GENERAL INFORMATION:
APPLICANT: KUNST, Frederik
TITLE OF INVENTION: Listeria innocua, genome and applications
FILE REFERENCE: 344 702 - US
CURRENT FILING DATE: 2003-03-27
PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
PRIOR FILING DATE: 2001-10-04
PRIOR APPLICATION NUMBER: FR 00/12 697
PRIOR FILING DATE: 2000-10-04
NUMBER OF SEQ ID NOS: 4025
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3563
LENGTH: 2751
TYPE: DNA
ORGANISM: Listeria monocytogenes 4b
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(end)
OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3563

Query Match 7.5%; Score 34.4; DB 15; Length 2751;
Best local similarity 52.0%; Pred. No. 1.1;
Matches 77; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 238 GATGATCCAGCGGTGGGATTAATGACTTGAAGCTGAGATGGTTTCAGAAT 297
DB 321 GATTAATGAAAGAGGAAATATTTACTTATGCAACAACGTTTCTGATATACATAT 380
QY 298 AATGCCACATCGACCGATGAAAGCTTAAAGCTGCATATTAATGTCGGCCATATGAT 357
DB 381 AAGAGCAATGATGATCTTTTGGCAGATTAACCAATATTCGTTTGGTTAAGAAATTA 440

OTHER INFORMATION: Clone ID: PAT_MRT3847_51326C.1
US-10-424-599-88945

		Indels	Gaps
QY	101 GCGCGGCAATAGTTCGGSCCGGACTCAAGCTTGAGCATTTATCGTACGGTTCGGTA	0	0
Db	556 GGGCCCAAGAAGGTATACGCCCTCGATACAGCGGGTCCATCTGTGATATATGTGTG	160	
QY	161 ACCGCTCGCTTGCTCTGCAAAAGCATGCCCGTAATCTGMAAGACACTTACCCAGACG	497	
Db	496 ATGATGTGATATATAGGATGATGAGCATGAGGAGGTGGCAACGAAGATGTAAACGGG	220	
QY	221 GTTATGCTAAGCGCGCGATGTAGGCCAGAGGTCCGATAT	437	
Db	436 GTTATGCTACTAGAGGTAAATGCAAGGAGAGTTCCGGTA	398	

407-13.rnpb

	5'	3'	Mismatches	55;	Indels	0;	Gaps	0;
Qy	52	GCTCTGCTGGCGCGTGGTTCCACATAGGAGCGCGCGCGGTATCATATAGGCGCGCGCAT						111
Db	209	GCTCCGGTTGCCGCACTACCTGATTTGGAGATAGAGATGACTTTAAAGTTCCGCCAAT						150
Qy	112	AGTTCCGCGCCCGGACTCAACTTATGACATTATACAGTTCGGCTTACAGGCTCGCGCTT						171
Db	149	AATTGAGCCATTGTGCTCAATGGTTTGGCCATCAGTGGGTGGTCCCAACGCAAAATTT						90
Qy	172	GC 173						
	89	GC 88						
Db								

US-09-880-371-6/C
Sequence 6, Application US/09880371
Patent No. US2002005958A1
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Derocher, Jay
TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
FILE REFERENCE: 21829/91
CURRENT APPLICATION NUMBER: US/09/880,371
CURRENT FILING DATE: 2001-06-13
PRIOR APPLICATION NUMBER: 60/211,585
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 6
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-880-371-6

	Indels	Gaps
QY 52 GCTCTGCTGGCGTGGTCCACATGGCGCGCGCGTAACTAAACGGCGCGGCAAT	55	0
Db 209 GCTCCCGGTGGCGCATTAACCTGATTTGGGAGAAAGACGATTTAAAGATCCGCCAAT		111
QY 112 AGTTCGGCGCCGGACTCAACGTTGAGCATTTATCAAGTACGGTCCGGCAACGCTGCGCTT		150
Db 149 AATTGAGCCATTTCCTAATGATTTCCGATCGATGGGTGTTCCCCCAAGGAAATTT		171
QY 172 GC 173		90
Db 89 GC 88		

US-09-879-248-5/c
; Sequence 5, Application US/09879248
; Patent No. US20020062500A1
; GENERAL INFORMATION:
; APPLICANT: Fan, Hao
; APPLICANT: Wei, Zhong-Min
; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
; FILE REFERENCE: 21829/81
; CURRENT APPLICATION NUMBER: US/09/879,248
; CURRENT FILING DATE: 2001-06-12
; PRIOR APPLICATION NUMBER: 60/212,211
; PRIOR FILING DATE: 2000-06-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.1

; GENERAL INFORMATION:

```

RESULT:10
US-10-156-761-1
: Sequence 1, Application US/10156761
: Publication No. US20030115018A1
GENERAL INFORMATION:
APPLICANT: OMDURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HAYTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249--262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

	Accession	U	Gaps	0
OY	157	GC	TAACGCTGCGCTTGGCTCTGTCGAAAGCGCATGCGCCGTTAAATCTGAAACGACCAATTACCCAG	216
Db	4432	GT	TACCGGCTGCGCGCTGCTGTGTGACAGAGATGAGAGAGAGCTTGCAAGAGGCTCAAAATCCG	4491
OY	217	AG	CGTATTATGTATGATGACGGCGCCGATGTGTGGCCAGGGTGCAGATTAATAGTACTTATGAACTG	276
Db	4492	AT	TGCTGCTGCTATTGGGAAAGGGTATGAGATGATGTGTGCGATTTTGTATGAGATGAGATG	4551
OY	277	ACT	CGAATGGTTTCAGAAATTAATGCCACCATGACACATGGAACGCTTAAAACTCCGAT	336
Db	4552	AAGA	AGATGATCATGATTTGCTGCTGCTATTTCCTTTGAGGGGATGGCTTACGATGAGAAAG	4611
OY	337	ATT	ACTGTGCGCCAAATTGAT	357
Db	4612	TTTG	CTGTGAAGACTAATGAT	4632

RESULT 12
US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.


```

; Sequence 24385, Application US/10289122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
;
; APPLICANT: Wang, Liangu
;
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert

```

```

; APPLICANT: Ohlsen, Kari
; APPLICANT: Zvekind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 24385
; LENGTH: 1776
; TYPE: DNA
; ORGANISM: Listeria monocytogenes
US-10-282-122A-24385

```

```

Query Match      7.2%; Score 32.8; DB 12; Length 1776;
Best Local Similarity 51.4%; Pred. No. 3;
Matches 76; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 238 GATGAGCCCGGGTGGCGATTAAGTACTTGAAGTGAAGTGGTTTCAGAAAT 297
DB 775 GATGTTGTCAAGAAAGCAAAAATATCACCATTGAAAAACGTTTCTCGTATGACATAT 834

QY 298 AATGCCACCATCGACCGAGTGAAGCTAATAAATCCGATATTACTGTGCGCAATATGAT 357
DB 835 AAAAGAGCAATGATCGTTTGGCAGTGAATAAACCGGATATTCGTTTGGTTAGAAATTA 894

QY 358 CAGCTGTTACCGCGTGTGTTTACCCATG 385
DB 895 CAAATGTATCCGATGCTGTAAAGATG 922

```

Search completed: March 17, 2004, 08:15:58
Job time : 415.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacagctatcaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_estba:*
3: em_estba:*
4: em_estba:*
5: em_estba:*
6: em_estba:*
7: em_estba:*
8: em_estba:*
9: em_estba:*
10: em_estba:*
11: em_estba:*
12: em_estba:*
13: em_estba:*
14: em_estba:*
15: em_estba:*
16: em_estba:*
17: em_estba:*
18: em_estba:*
19: em_estba:*
20: em_estba:*
21: em_estba:*
22: em_estba:*
23: em_estba:*
24: em_estba:*
25: em_estba:*
26: em_estba:*
27: em_estba:*
28: em_estba:*
29: em_estba:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	226.2	49.6	680	12	Bu618688
C 2	36	7.9	523	12	Bu334624
C 3	35.6	7.8	744	14	CF446143
C 4	34.8	7.6	549	28	B2295150

Result No.	Score	Query Match	Length	ID	Description
C 5	34.8	7.6	594	12	Bu335653
C 6	34.8	7.6	813	28	B2204853
C 7	34.6	7.6	540	14	CB210843
C 8	34.6	7.6	556	9	AV976624
C 9	34.4	7.5	539	13	BQ141266
C 10	34.4	7.5	560	10	AW686303
C 11	34.4	7.5	627	10	AW573882
C 12	34.2	7.5	456	12	Bu331339
C 13	34.2	7.5	524	12	Bu367372
C 14	34.2	7.5	547	12	Bu333958
C 15	34.2	7.5	553	12	Bu336903
C 16	34.2	7.5	569	12	Bu334047
C 17	34.2	7.5	574	12	Bu330328
C 18	34.2	7.5	607	12	Bu355766
C 19	34.2	7.5	625	12	Bu356922
C 20	34.2	7.5	629	12	Bu335378
C 21	34.2	7.5	635	12	Bu397391
C 22	34.2	7.5	645	12	Bu392753
C 23	34.2	7.5	670	12	Bu336637
C 24	34.2	7.5	1036	12	BG671547
C 25	34	7.5	234	10	BB581987
C 26	34	7.5	500	14	CA711477
C 27	34	7.5	512	10	BB757100
C 28	33.8	7.4	298	14	F19524
C 29	33.8	7.4	365	9	AA746477
C 30	33.8	7.4	558	14	CD374421
C 31	33.8	7.4	559	13	Bu765156
C 32	33.8	7.4	941	29	CG178280
C 33	33.8	7.4	1098	29	CNS02GQP
C 34	33.6	7.4	712	14	CF437982
C 35	33.6	7.4	811	28	BZ469000
C 36	33.4	7.3	402	13	Bu767674
C 37	33.4	7.3	505	6	AL809250
C 38	33.4	7.3	536	13	Bu767674
C 39	33.4	7.3	546	13	Bu780674
C 40	33.4	7.3	699	12	B1959328
C 41	33.4	7.3	715	13	Bu444683
C 42	33.2	7.3	589	28	BZ300068
C 43	33.2	7.3	629	9	AU295325
C 44	33.2	7.3	788	9	AU139321
C 45	33.2	7.3	1646	10	BF579120

ALIGNMENTS

RESULT 1
Bu618688/c
LOCUS
DEFINITION
Bu618688 NIBB Mochii normalized Xenopus early gastrula library
Bu618688
Xenopus laevis CDNA clone XL186b22 5', mRNA sequence.

ACCESSION
Bu618688
VERSION
Bu618688.1 GI:37256713

KEYWORDS
EST.

SOURCE
Xenopus laevis (African clawed frog)

ORGANISM
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.

REFERENCE
1 (bases 1 to 680)
Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-i, T. and

AUTHORS
Kohara, Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
The information of this clone is available through the following URL.

TITLE
JOURNAL
COMMENT

FEATURES
source
http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. .680

/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="XL18b22"
/issue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 49.6%; Score 226.2; DB 12; Length 680;
Best Local Similarity 68.4%; Pred.No.3,5e-53;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGCGACATTCGACCAATCGATGTTTCTGCGAGTGGCTGGCT 60
DB |||||||
593 ATGAACCTTTTAAAGTGCGACATTCGACCAATCGATGTTTCTGCGAGTGGCTGGCT 60
QY 61 GCGCTGCTTCACAAATGGGCGGCGGCTATCATTAAGCGCGGCAATAGTTCGGC 120
DB |||||||
533 GGTGTTGTTCTCAGTACGCGCGGCGGTAACCGAGTGGTGGCGGTATATAGCGGC 474
QY 121 CCGGACTCAAGTTGAGCATTTATCAGTAGCGGTTCCGCTAACGCTGGCTGCTGCA 180
DB |||||||
473 CCAATTTTGAAGTGAACATTTACAGTACGCGGCGGTAACCTGCTGCA 414
QY 181 AGCGATGCGCGTAAATCGAAAGCAATCCCAAGCGGTTATGTAACCGCGCGCAT 240
DB |||||||
413 ACTGATGCGCGTAACTCTGACTGATTAATCCCAAGCGGCGGTTATGTAACCGCAT 354
QY 241 GTAGCGCGGTTGGGTAATAGTACTATTAAGTACTGATGATGATGTTTCAATAT 300
DB |||||||
353 GTTGTGAGGCGTCAAGTACAGCTCAATGATGATGATGATGATGATGATGATGAT 294
QY 301 GCCACCATCGACCGTGAAGCGTAAAGTCCGATATTAATGTTGCGGCAATATGATCAG 360
DB |||||||
293 GCTACTCTGATAGTGAAGCGCAAAATTTCTGAATGAGCGTTAAACAGTTGGTGT 234
QY 361 CTGCTTACCGCTGTTTGAATCCATGAATGAGCAATGCAAGTAAATGTTGCGGCAAT 420
DB |||||||
233 GCGACGCGTCTGAGTGTGACGACGATGATTAATCTCTCGTCAAGTGTACAGTT 174
QY 421 GGTGTTGCGCAACGCGCGCTAACGATTTAA 456
DB |||||||
173 GGTGTTGTAACACGCGACCGCTCATGATGATTA 138

RESULT 2
BU334624 523 bp mRNA linear EST 05-MAR-2002
LOCUS BU334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
DEFINITION BU334624 Dictyostelium discoideum cDNA library, AF Dictyostelium
ACCESSION BU334624
VERSION BU334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
AUTHORS Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
REFERENCE 1 (bases 1 to 523)
TITLE Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-I,T.
Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehini@genes.nig.ac.jp.

FEATURES
source

Location/Qualifiers
1. .523
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda4706"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Query Match 7.9%; Score 36; DB 12; Length 523;
Best Local Similarity 55.5%; Pred.No.30;
Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 238 GATGAGCGGAGGTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 297
DB |||||||
398 GATGAGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 457
QY 298 AATGCCACATCGACCGATGAAACGCTAAACCTCCGATTAATGATTAATGATTAATGAT 356
DB |||||||
458 AATTAACAACAACAACAACAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 516

RESULT 3

CF446143 744 bp mRNA linear EST 04-SEP-2003
LOCUS CF446143 EST682488 normalized cDNA library of onion Allium cepa cDNA clone
DEFINITION ACAHU74, mRNA sequence.
ACCESSION CF446143
VERSION CF446143.1 GI:34468845
KEYWORDS EST.
SOURCE Allium cepa (onion)
ORGANISM Allium cepa

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Asparagales; Alliaceae;
Allium.

1 (bases 1 to 744)
Haver,M.J., Cheng,F., Van Aken,S., Uterback,T. and Town,C.D.

Expressed Sequence Tags from a normalized library of mixed onion
tissues (Allium cepa)
JOURNAL Unpublished (2003)
COMMENT Contact: Haver MJ
Department of Horticulture
USDA-ARS and University of Wisconsin
1575 Linden Drive, Madison, WI 53706, USA
Tel: 608-262-1830
Fax: 608-262-4743
Email: mjhavey@facstaff.wisc.edu
TIGR sequence name ACAHU74TR. For more information:
http://haveylab.hort.wisc.edu
Seq primer: CAG GAA ACA GCT ATG ACC.

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers
1. .744
/organism="Allium cepa"
/mol_type="mRNA"
/cultivar="Red Creole (bulbs), unknown (callus), Epano &
Texas legend (roots)"
/db_xref="taxon:4679"
/clone="ACAHU74"
/issue_type="Callus, roots, and young bulbs"
/clone_lib="normalized cDNA library of onion"
/note=Vector: pCMVSPORT6.1-ccdb (Invitrogen); Site 1:
from callus, roots, and young bulbs were combined to
synthesize the library. Normalization was performed by
low-copy transcripts was performed by proprietary
techniques of Invitrogen."

ORIGIN

Query Match 7.8%; Score 35.6; DB 14; Length 744;
Best Local Similarity 47.7%; Pred.No.46;

	Matches	104; Conservative	0; Mismatches	114; Indels	0; Gaps	0
QY	224	CGCCGATGTAGCCAGGGTGCGGATTAATATGTACTTTGGACTGACTCGAATAGTTTCAG	293			
Db	382	CGAGCTAAGAACCAACAACTTTAATAGCAATTTATCATCTGCATGTGATATGATATCG	323			
QY	294	AAATATATGCCACATCGACAGTGGAAAGCTAAATACTCCGATTTATCTGTGGGCCAATTA	353			
Db	322	AAATTAACATATGTAACAGAAATGGAATCTTTAAAGTGCTGTGATATCATTAATATGA	263			
QY	354	TGATAGCTGTTACCCGTGTTGTTACCCATGAAATGGCAATGCAAGCATATATGTGCG	413			
Db	262	AGAGGGAGAGAAACCAACTTTTATATGATGAATCTGTCTCAATATGTTGGCAAGAGAGG	203			
QY	414	TCAGGTGTGTTTGGCAACAACGCCACGGCTAACAGT	451			
Db	202	AAGATTTAACACTTCACACACAAATCGGCTTCGAAT	165			

RESULT 4					
B2295150/c	B2295150	549 bp	DNA	linear	GSS 31-OCT-2007
LOCUS	CG3320..r1	Candida glabrata Random Genomic Library Candida glabrata genomic clone CG3320, genomic survey sequence.			
ACCESSION	B2295150				
VERSION	B2295150.1	GI:24356228			
KEYWORDS	GSS,				
SOURCE	Candida glabrata				
ORGANISM	Candida glabrata				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
	Saccharomycetales; Microsporici Saccharomycetales; Candida.				
REFERENCE	1 (bases 1 to 549)				
AUTHORS	Wong, S., Fares, M.A., Zimmermann, W., Butler, G. and Wolfe, K.H.				
TITLE	Evidence from comparative genomics for a complete sexual cycle in the 'asexual' pathogenic yeast <i>Candida glabrata</i>				
JOURNAL	Genome Biol.	4 (2),	R10	(2003)	
MEDLINE	22508158				
PUBMED	12520120				
COMMENT	Contact: Wong S				

Department of Genetics, Smurfit Institute
Trinity College Dublin
Dublin 2, Ireland
Tel: 353 1 6082319
Tel: 353 1 6798558
Email: swong@tcd.ie
Class: plasma ends
Location/Qualifiers
1..549

```

/organism="Candida glabrata"
/mol_type="genomic DNA"
/strain="CBS_138"
/db_xref="taxon:5478"
/clone="CG1320"
/clone_lib="Candida glabrata Random Genomic Library"

```

Query Match	Similarity	Score	DB	Length
Best Local	52.0%	Pred. No. 68		549
Matches	78	Conservative	0	Mismatches 72; Indels 0; Gaps 0;
Qy	222	TTATGTAACGGCGCCGATGTAGGCCAGGCTGCCGATATAGTACTATGTGAACGACTCA	281	
Db	291	TTATAGTTTTTGTCTCTACAAAGGCCAAATTTTCAGTAAAAAGCTGAATATTTTGCA	232	
Qy	282	GAATGGTTTCGAAATTAATGCCACCATGACGACGTGGAAAGCTAAAACTCGATATTAC	341	
Db	231	GAATAATGACAGTGTGTATACGAAACACATCTAGATCGCAACCACTCTATATCAG	172	
Qy	342	TGTGGGCCAATATGATGACGTGTTACCG	371	
Db	171	TGTCGCCAAGAAATTCCTTTGATTAACCG	142	

RESULT 5					
LOCUS	BJ335653				
DEFINITION	BJ335653	594 bp	mRNA	linear	EST 05-MAR-2002
ACCESSION	BJ335653	Dictyostelium discoideum cDNA library, AF	Dictyostelium		
VERSION	BJ335653	dictoem cDNA clone dda51h06.5', mRNA sequence.			
KEYWORDS	BJ335653.1	GI:19165783			
SOURCE	Est.				
ORGANISM	Dictyostelium discoideum				
REFERENCE	Dictyostelium discoideum				
AUTHORS	Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.				
TITLE	1 (bases 1 to 594)				
JOURNAL	Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.				
COMMENT	Full length cDNA of Dictyostelium discoideum at the aggregation stage				
	unpublished (2002)				
	Contact: Tadasu Shin-i				

Center for Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel.: 81-559-81-6856
Fax: 81-559-81-6855
Email: tehin@genes.nig.ac.jp.
Location/Qualifiers
1. .594

```

/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda51h06"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_id="Dictyostelium discoideum cDNA library, Af"

```

Query Match	7.6%	Score	34.8	DB	12	length	594
Best Local Similarity	55.5%	Pred. No.	70				
Matches	66	Conservative	0	Mismatches	53	Indels	0
						Gaps	0

[illegible]

RESULT 6	BZ204853	813 bp	DNA	linear	GSS 11-OCT-2002
LOCUS	BZ204853				
DEFINITION	BZ204853				
CH230-390D21.TV CHORI-230 Segment 2					
CH230-390D21, genomic survey sequence.					
ACCSSION	BZ204853				
VERSION	BZ204853.1	GI:23862905			

SOURCE ORGANISM	REFERENCE AUTHORS	TITLE JOURNAL COMMENT
Rattus norvegicus (Norway rat)		
Rattus norvegicus		
Eukaryotes; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
1 (bases 1 to 813)		
Zhao, S., Shetty, J., Shatman, S., Tesgaye, G., Geer, K., Shvartsbeyn, A., Gegregory, E., Overton, L., Russell, D., Chen, D., Riggs, F., De Jong, P. and Fraser, C.M.		
Rat BAC End Sequences from Library CHORI-230 MboI segment Unpublished (1999)		
Other GSSs: CH230-390D21.TJ		
Contact: Shaying Zhao		
Department of Eukaryotic Genomics		
The Institute for Genomic Research		
9712 Medical Center Dr., Rockville, MD 20850, USA		
Tel.: 301 838 0200		

Fax: 301 838 0208
 Email: szhae@tigr.org
 Clones are derived from the rat BAC library CHORI-230
 (http://www.chori.org/bacpac/rat230.htm). For BAC library
 availability, please contact Pieter de Jong (pdejong@mail.cho.org).
 Clones may be purchased from BACPAC Resources
 (http://www.chori.org/bacpac/orering/information.htm). BAC end
 page: http://www.tigr.org/tdb/bac_end/rat/bac_end_intro.html
 Plate: 390 row: D column: 21
 Seq primer: 77
 Class: BAC ends

FEATURES

source
 1. 813
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /strain="BN/SNHsd/MCM"
 /db_xref="taxon:10116"
 /clone="CH230-390D21"
 /sex="Female"
 /cell_type="Brain"
 /clone_lib="CHORI-230 Segment 2"
 /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
 CHORI-230 Rat (BN/SNHsd/MCM) BAC library produced by
 Pieter de Jong"

ORIGIN

Query Match 7.6%; Score 34.8; DB 28; Length 813;
 Best Local Similarity 50.0%; Pred. No. 80;
 Matches 87; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 Oy 178 CAAAGCGATGCCCTTAATCTGAACGACCATTAACCGGGTATGTAGCGGCC 237
 Db 548 CAAAGCGATGCCCTTAATCTGAACGACCATTAACCGGGTATGTAGCGGCC 237
 Oy 238 GATGTAGCGCGGGTGGGATATGATCTATTGAACTGCTAGATGGTTTGAAGAT 297
 Db 608 TCTGGAACATGATGGGACCTGGAATAAATTCCTGGAACAAACACCAATGGCTTATGCTCT 667
 Oy 298 AATGCCCATGACGACGATGGAACGCTAAATCTCGGATATGCTGTGGCCAA 351
 Db 668 AAGATCAAGATCGACCAATGGATCTCAATTAAGCTTCTGTAGGCAAA 721

RESULT 7 540 bp mRNA linear EST 05-FEB-2003
 LOCUS CB210843
 DEFINITION OML01123 Oryza minuta HybridZAP-2.1 XR library Oryza minuta cDNA 5',
 mRNA sequence.
 ACCESSION CB210843
 VERSION CB210843.1 GI:28256934
 KEYWORDS EST.
 SOURCE Oryza minuta
 ORGANISM Oryza minuta
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 1 (bases 1 to 540)
 Shin, Y.S.
 Oryza minuta HybridZAP-2.1 XR library
 Unpublished (2003)
 Contact: Jeong Sheop Shin
 Plant Molecular Genetics
 Graduate School of Biotechnology, University of Korea
 136-701 Anam-dong 5/1 Seoul, Korea
 Tel: 00 82 2 3290 3430
 Fax: 00 82 2 927 9028
 Email: jsheo@kucn.korea.ac.kr.

FEATURES
 source
 1. 540
 Location/Qualifiers
 /organism="Oryza minuta"
 /mol_type="mRNA"
 /db_xref="taxon:63629"
 /dev_stage="4-weeks after germination"

/clone_lib="Oryza minuta HybridZAP-2.1 XR library"
 /note="Organ: Immature leaf"

ORIGIN

Query Match 7.6%; Score 34.6; DB 14; Length 540;
 Best Local Similarity 47.1%; Pred. No. 77;
 Matches 85; Conservative 0; Mismatches 84; Indels 0; Gaps 0;
 Oy 225 TGTAAAGCGCCGATGATGAGCCAGGCTGGGATTAATGACTATGAACTGACTCAGA 284
 Db 140 TGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 199
 Oy 285 TGTGTTTCAAGAAATATGACCAATGACGACGATGGAACGCTAAATCTCGATATTAAGT 344
 Db 200 TGTGTTTCAAGAAATGCAATGACGACGATGGAACGCTGGAATCAATTAATCTTAT 259
 Oy 345 CGGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 393
 Db 260 GAAAGCTCTCCAAAGACAGATGATGATGATGATGATGATGATGATGATGATGATG 308

RESULT 8

AV976624

AV976624 Nori Satoh unpublished cDNA library, egg Clona EST 14-MAR-2002

intestinalis cDNA clone cleg42j24 5', mRNA sequence.

AV976624

AV976624.1 GI:19466390

EST.

Clona intestinalis

Clona intestinalis

Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;

Plebobranchia; Clonidae; Clona.

1 (bases 1 to 556)

Satoh, N., Satoh, Y., Kohara, Y. and Shin, I. T.

Expressed genes in Clona intestinalis

Unpublished (2000)

Contact: Nori Satoh

Department of Zoology

Kyoto University

Sakyo-ku, Kyoto, Kyoto 606-8502, Japan

Tel: 81-75-753-4081

Fax: 81-75-705-1113

Email: satoh@ascidian.zool.kyoto-u.ac.jp.

Location/Qualifiers

1. 556

/organism="Clona intestinalis"

/mol_type="mRNA"

/db_xref="taxon:7719"

/clone="cleg42j24"

/tissue_type="whole animal"

/dev_stage="egg"

/clone_lib="Nori Satoh unpublished cDNA library, egg"

ORIGIN

Query Match 7.6%; Score 34.6; DB 9; Length 556;
 Best Local Similarity 47.1%; Pred. No. 78;
 Matches 106; Conservative 0; Mismatches 119; Indels 0; Gaps 0;
 Oy 222 TTATGTTAAGCGCCGATGATGAGCCAGGCTGGGATTAATGACTATGAACTGACTCA 281
 Db 67 TGAATCAAGCTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 126
 Oy 282 GAATGTTTCAAGAAATATGACCAATGACGACGATGGAACGCTAAATCTCGATATTAAG 341
 Db 127 TGAATGACCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 186
 Oy 342 TGTGCGCAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 401
 Db 187 TGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246
 Oy 402 CGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 446

Db 247 TGATGTCAAGCTGATGATGCTACTAATGATGTACACCTAA 291

RESULT 9
LOCUS B0141266
DEFINITION B0141266 539 bp mRNA linear EST 26-APR-2002
NP017603PH1024 Phoma-infected Medicago truncatula cDNA clone
B0141266
ACCESSION B0141266
VERSION B0141266
KEYWORDS B0141266.1 GI:20277392
SOURCE EST
ORGANISM Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 539)
REFERENCE Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
AUTHORS Gonzales,R.A., Bell,C.J., Inman,J.T., Maugh,M.E., Sullivan,J.P.,
May,G.D. and Paiva,N.L.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula Phoma-infected library
COMMENT Unpublished (2002)
Contact: Paiva NL
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317
Fax: 580 221 7380
Email: nlpaiva@noble.org
Insert Length: 539 Std Error: 0.00
Plate: 017 row: G column: 03
Seq primer: TCACACAGAAACGCTATGAC.
Location/Qualifiers
1. 539
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF017603PH"
/tissue_type="leaf"
/dev_stage="Pathogen-induced, young trifoliolate"
/note="Vector: pluescript SK(-); Young trifoliolate leaves
of Medicago truncatula were excised and dip-inoculated in
a spore suspension of Phoma medicaginis, and incubated in
humid dishes. Pools of leaves were harvested at 0, 15, and
30 minutes and 1, 2, 3, 6, 14, 24, 48, 72, and 96 hours,
and used to prepare total RNA. cDNA was prepared from
polyA+ enriched, pooled samples of equivalent amounts of
total RNA from each sample. The cDNA was directionally
ligated into the Uni-Zap XR vector (Stratagene) and
packaged using the Gigapack III Gold packaging extracts.
Phagemids containing cDNA inserts were in vivo excised
from the recombinant Uni-ZAP XR vector using EXSist
helper phage and the E. coli strain XL1-Blue MRF'
(Stratagene). Excised plasmids were plated using SOLR
cells."

ORIGIN
Query Match 7.5%; Score 34.4; DB 13; Length 539;
Best Local Similarity 46.0%; Pred. No. 88;
Matches 116; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 129 AACCTGAGCATTTATAGTAGCGTTCCGCTAACGTCGCTCTGCAAGCGATGC 188
DB 233 AATTTTAAAAACATCTACTCCGGTTACAATAATGCAATCCCTTGAATGAAGACGAGGC 292
QY 189 CCGTAATTCGAAACGACCATTAACCCAGAGCGGTTATGTAACGCGCCGATGAGGCCA 248
DB 293 ACGTGATCCACCTGAGATTCCTCCACCAATGCAAAACCTTGTGGCTATCTCTGCTAA 352
QY 249 GGGTGGCATTAATAGTACTATTAAGAACTGACTCAGAAATGTTTCAGAAATAATGACACCAT 308

Db 353 TGTGATAGTTTATGTTCTCTTCATTCCTCCAGACGCAATTCMAATCATCTTAAAT 412

QY 309 CGACCATGGAACCGCTAAACATCCGATATTACTGTGCGCCAAATATGATCAGCTGTTAC 368
DB 413 TGAGAAATGAGAGGCACAAAGATCTGTGTGGCTTTGGGATACACTGCTTCTGTAC 472
QY 369 CCGTGTGTTAC 380
DB 473 CAAATTTGTTAC 484

RESULT 10
LOCUS AM686303
DEFINITION AM686303 560 bp mRNA linear EST 20-DEC-2000
NP036006NR1000 Nodulated root Medicago truncatula cDNA clone
AM686303
ACCESSION AM686303
VERSION AM686303
KEYWORDS AM686303.2 GI:11931130
SOURCE EST
ORGANISM Medicago truncatula (barrel medic)
Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 560)
REFERENCE Watson,B.S., Shin,H.-S., Lopez-Meyer,M., Scott,A.D., Harris,A.R.,
AUTHORS Gonzales,R.A., Bell,C.J., Flores,H.R., Inman,J.T., Weller,J.W.,
May,G.D. and Paiva,N.L.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
JOURNAL Medicago truncatula nodulated root library
COMMENT Unpublished (2000)
On Apr 14, 2000 this sequence version replaced gi:7561039.
Contact: Paiva NL
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7317
Fax: 580 221 7380
Email: nlpaiva@noble.org
Insert Length: 665 Std Error: 0.00
Plate: 036 row: D column: 06
Seq primer: TCACACAGAAACGCTATGAC.
Location/Qualifiers
1. 560
/organism="Medicago truncatula"
/mol_type="mRNA"
/db_xref="taxon:3880"
/clone="NF036D06NR"
/tissue_type="root"
/dev_stage="Pooled developmental"
/note="Vector: Lambda Zap; Four-week old Rhizobium
meliloti-inoculated Medicago truncatula roots, containing
a mixture of young and old roots and nodules."

ORIGIN
Query Match 7.5%; Score 34.4; DB 10; Length 560;
Best Local Similarity 46.0%; Pred. No. 89;
Matches 116; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 129 AACCTGAGCATTTATAGTAGCGTTCCGCTAACGTCGCTCTGCAAGCGATGC 188
DB 230 AATTTTAAAAACATCTACTCCGGTTACAATAATGCAATCCCTTGAATGAAGACGAGGC 289
QY 189 CCGTAATTCGAAACGACCATTAACCCAGAGCGGTTATGTAACGCGCCGATGAGGCCA 248
DB 290 ACGTGATCCACCTGAGATTCCTCCACCAATGCAAAACCTTGTGGCTATCTCTGCTAA 349
QY 249 GGGTGGCATTAATAGTACTATTAAGAACTGACTCAGAAATGTTTCAGAAATAATGACACCAT 308
DB 350 TGTGATAGTTTATGTTCTCTTCATTCCTCCAGACGCAATTCMAATCATCTTAAAT 409

[illegible]

KEYWORDS EST. Dictyostelium discoideum
 SOURCE Dictyostelium discoideum
 ORGANISM Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 REFERENCE 1 (bases 1 to 524)
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadaeu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..524
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="ddc42p08"
 /sex="mat A"
 /dev_stage="Culmination stage"
 /clone_lib="Dictyostelium discoideum cDNA library, CP"
 ORIGIN
 Query Match 7.5%; Score 34.2; DB 12; Length 524;
 Best Local Similarity 55.5%; Pred. No. 99;
 Matches 66; Conservative 0; Mismatches 53; Indels 0; Gaps 0;
 Oy 238 GATGTAGCCACGAGGTGCGATTAATGTACTATTGAAGTCACTCAGATGTTTCAGAAAT 297
 Db 181 GATGAAGATGATGATTAATGAGAAATGATATATTGGAATCAATAGAAATGAGGACATATAT 240
 Oy 298 AATGCCACCATGACCACTGACATGCAACGCTTAAACTCCGATATTACTGTGGCCCATATGA 356
 Db 241 AATAACACCAACACACATTAATTAATAAATAATTAATTAATTAAGATGATGATGA 299
 RESULT 14
 LOCUS BJ333958 547 bp mRNA linear EST 05-MAR-2002
 DEFINITION BJ333958 Dictyostelium discoideum cDNA library, AF Dictyostelium
 accession cDNA clone dda44j16 5', mRNA sequence.
 VERSION BJ333958
 KEYWORDS BJ333958.1 GI:19164088
 SOURCE EST.
 ORGANISM Dictyostelium discoideum
 SOURCE Dictyostelium discoideum
 REFERENCE Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 AUTHORS 1 (bases 1 to 547)
 Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadaeu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..547
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dda44j16"
 /sex="mat A"
 /dev_stage="Aggregation stage"

```

ORIGINAL                               /clone_lib="Dictyostelium discoideum cdna library, Af"

Query Match      7.5%; Score 34.2; DB 12; Length 547;
Best Local Similarity .55.5%; Pred. No. 1e+02;
Matches        66; Conservative    0; Mismatches   53; Indels     0; Gaps     0;

QY      238 GATGTAGGCCGCGGTGGCGAATAATAGTACTATTGAACGCTCAGAATGTTTCAGAAAT 297
           |||||          |||||          |||||          |||||          |||||
DB      188 GATGAAGATGATGAATTGGAGAAATAGATATTTTGCAATCATAGAAAATGAGACAATTAAT 247

QY      238 AATGCCACCATGCAGCCAGTGGAGCGCTAAAAACTCGCATATTACTGTGCGCCAATATGA 356
           |||||          |||||          |||||          |||||          |||||
DB      244 AATAACAACAACAACAACAATTAATAATAAAAAATAAATAATTAATAAAGATGATGATGA 302


RESULT 15
LOCUS       BJ336903                    553 bp            mRNA             linear         EST 05-MAR-2002
DEFINITION  BJ336903 Dictyostelium discoideum cdna library, Af dictyostelium
ACCESSION   BJ336903
VERSION     BJ336903
KEYWORDS    Dictyostelium discoideum
SOURCE      Dictyostelium discoideum
ORGANISM    Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
REFERENCE   1 (bases 1 to 553)
AUTHORS     Urushihara,H., Tanaka,Y., Kohara,Y. and Shii-n,T.
TITLE       Full length cDNA of Dictyostelium discoideum at the aggregation stage
JOURNAL     Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
              Center For Genetic Resource Information
              National Institute of Genetics
              111 Yata, Mishima, Shizuoka 411-8540, Japan
              Tel: 81-559-81-6856
              Fax: 81-559-81-6855
              Email: tehin@genes.nig.ac.jp.
FEATURES             location/Qualifiers
     source          1..553
                     /organism="Dictyostelium discoideum"
                     /mol_type="mRNA"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddas5g13"
                     /sex="mat A"
                     /dev_stage="Aggregation stage"
                     /clone_lib="Dictyostelium discoideum cdna library, Af"

ORIGIN

Query Match      7.5%; Score 34.2; DB 12; Length 553;
Best Local Similarity .55.5%; Pred. No. 1e+02;
Matches        66; Conservative    0; Mismatches   53; Indels     0; Gaps     0;

QY      238 GATGTAGGCCGCGGTGGCGAATAATAGTACTATTGAACGCTCAGAATGTTTCAGAAAT 297
           |||||          |||||          |||||          |||||          |||||
DB      184 GATGAAGATGATGAATTGGAGAAATAGATATTTTGCAATCATAGAAAATGAGACAATTAAT 243

QY      298 AATGCCACCATGCAGCCAGTGGAGCGCTAAAAACTCGCATATTACTGTGCGCCAATATGA 356
           |||||          |||||          |||||          |||||          |||||
DB      244 AATAACAACAACAACAACAATTAATAATAAAAAATAAATAATTAATAAAGATGATGATGA 302


Search completed: March 16, 2004, 04:29:05
Job time : 2235.91 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds
(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgctaccagctatata 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenBml:.*
1: gb_ba:.*
2: gb_hcg:.*
3: gb_in:.*
4: gb_om:.*
5: gb_ov:.*
6: gb_pat:.*
7: gb_ph:.*
8: gb_pl:.*
9: gb_pr:.*
10: gb_ro:.*
11: gb_str:.*
12: gb_sy:.*
13: gb_vl:.*
14: gb_vl:.*
15: em_ba:.*
16: em_fun:.*
17: em_hum:.*
18: em_in:.*
19: em_mu:.*
20: em_om:.*
21: em_or:.*
22: em_ov:.*
23: em_pat:.*
24: em_ph:.*
25: em_pl:.*
26: em_ro:.*
27: em_str:.*
28: em_un:.*
29: em_vl:.*
30: em_hcg_hum:.*
31: em_hcg_inv:.*
32: em_hcg_other:.*
33: em_hcg_mus:.*
34: em_hcg_pln:.*
35: em_hcg_rtd:.*
36: em_hcg_mam:.*
37: em_hcg_vrt:.*
38: em_ey:.*
39: em_htgo_hum:.*
40: em_htgo_mus:.*
41: em_htgo_other:.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403.2	88.4	456	6	144909
2	403.2	88.4	2067	1	SEU43280
3	395.2	86.7	5103	1	STRAJ2301
4	395.2	86.7	22411	1	AE008749
5	393.6	86.3	254050	1	AL627269
6	393.6	86.3	301983	1	AE016840
7	387.8	85.0	1048	1	STAFBFA
8	284.2	62.3	361	6	144908
9	278.4	61.1	2889	1	CSP515700
10	229.2	50.3	2920	1	CPS515701
11	225.4	49.4	1711	1	AF275733
12	225.4	49.4	10190	1	AE005315
13	225.4	49.4	327773	1	AE002554
14	224	49.1	4680	1	ECCSGABDG
15	224	49.1	10346	1	AE000205
16	224	49.1	15047	1	D90741
17	223.8	49.1	306358	1	AE016759
18	222.4	48.8	456	6	AX814811
19	222.4	48.8	648	1	ECOCGAA
20	208	45.6	2883	1	ESAS15702
21	203.2	44.6	230	1	SEU53207
22	186.4	40.9	10370	1	AE015131
23	186.4	40.9	292504	1	AE016981
24	163.4	35.8	437	1	AF237726
25	93.2	20.4	19201	1	D90742
26	50	11.0	1212	1	EC0131756
27	48.2	10.6	78	6	AX814809
28	41.6	9.1	72	6	AX814798
29	36.2	7.9	1425	1	AF343445
30	36.2	7.9	302156	1	AC116977
31	35.6	7.8	13430	1	AE000409
32	35.6	7.8	108366	2	AC020660
33	35.6	7.8	110000	1	ECOWM67_2
34	35.6	7.8	110000	2	AC145807_0
35	35.6	7.8	196539	2	AC145912
36	35.4	7.8	168267	2	AC146009
37	35.2	7.7	163619	2	AC118493
38	34.6	7.6	11160	1	AE008209
39	34.6	7.6	11476	1	AE009405
40	34.6	7.6	125020	9	AF429315
41	34.2	7.5	1458	1	AF173043
42	34.2	7.5	179906	2	AC119683
43	34.2	7.5	268098	2	AC133387
44	34.2	7.5	299708	2	AC110963
45	34	7.5	956	8	AK058943

ALIGNMENTS

RESULT 1
LOCUS 144909 456 bp DNA
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION 144909
VERSION 144909.1 GI:2469622
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Dorman,J.V., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

Pred. No. is the number of results predicted by chance to have a

FEATURES

Location/Qualifiers

1..456
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match

Best Local Similarity 88.4%; Score 403.2; DB 6; Length 456;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGGCGCTCTGGCT 60
DB 1 ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTAGTTCTGGCGCTCTGGCT 60
QY 61 GCGCTTATGATCAGCTGTTACCCGTTGTTACCCATGAATGACATGATCCGGC 60
DB 61 GCGCTTATGATCAGCTGTTACCCGTTGTTACCCATGAATGACATGATCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCGATACGGTCCGCTAACGCTGCTGCTGCAA 120
DB 121 CCGGACTCAACGTTGAGCATTTATCGATACGGTCCGCTAACGCTGCTGCTGCAA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGCTAAACGCGCGAT 180
DB 181 AGCGATGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGCTAAACGCGCGAT 240
QY 241 GTAGGCGAGGTCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
DB 241 GTAGGCGAGGTCGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GGCACCATTCGACCACTGGAACGCTAAACCTCCGATTAATCTGTCGCGCAATACGCGGT 300
DB 301 GGCACCATTCGACCACTGGAACGCTAAACCTCCGATTAATCTGTCGCGCAATACGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGGTAATGTAATGTAAT 360
DB 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGGTAATGTAATGTAAT 420
QY 421 GGTTCGCAACACGCGGCTTAACAGATTTAA 456
DB 421 GGTTCGCAACACGCGGCTTAACAGATTTAA 456

RESULT 2
LOCUS SEU43280 2067 bp DNA linear BCT 14-FEB-1996
DEFINITION *Salmonella enteritidis* agfBAC operon: fimbriin-like protein
precursor (agfB), thin aggregative fimbriae precursor (agfA), and
AgfC (agfC) genes, complete cds.
VERSION U43280.1 GI:1184712
KEYWORDS

SOURCE

ORGANISM *Salmonella enteritidis*
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; *Salmonella*.
1 (bases 1 to 2067)
Doran, J.L., Collinson, S.K., Burian, J., Santos, G., Todd, B.C.,
Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I., and Kay, W.W.
DNA-based diagnostic tests for *Salmonella* species targeting agfA,
the structural gene for thin, aggregative fimbriae
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
94013373
8104955
2 (bases 1 to 2067)
Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
Kay, W.W.
Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae
J. Bacteriol. 178 (3), 662-667 (1996)
96146512
8550497
3 (bases 1 to 2067)

REFERENCE

JOURNAL MEDLINE PUBLISHED
TITLE
AUTHORS
REFERENCE

AUTHORS

Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
Direct Submission
Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
Victoria, BC V8W 3P6, Canada
Location/Qualifiers
1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrc"
571..576
598..603
696..1151
/gene="agfB"
696..1151
/gene="agfB"
/note="fimbriin-like protein precursor"
/codon_start=1
/transl_table=11
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKKKLIFMMLTIIGAPGATATNYDLASEYNFAVNELSKSFN
QALITGVGTNSARVROEGSKLSVDSOEGNNRAKVDQAGNPFAYIEQGNANDA
SISQAYGNSAIIQKSGNKAMITQGTOKTAVVVKOSHMAIRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal
fimbriin-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbriin subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=11
/product="AgfA fimbriin precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLKVAFPAIVVSSALAGVVPWGCGGNGGNSGSPDST
LSIYVGSANALALQSDARKSEPTTQSGTNGADVGADNSTIELTONGFRNAT
IDWNAKNSDITVQYGNNAALVNOTASDSVVRQVFGNNATANGY"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/product="AgfA fimbriin"
/note="thin aggregative fimbriin subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=11
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MHTILALASNGITFTTQGGDIYVLPQVTLNRPVCQVOLT
SYRDSGVGSHRQOKQTLSPANQPIELSYNISSESDSVKILTVSDGSLHLSQO
WPSAQ"

FEATURES

source

-35_signal
-10_signal
gene
CDS

sig_peptide
mat_peptide

gene
CDS

sig_peptide
mat_peptide

stem_loop
gene
CDS

ORIGIN

Query Match 88.4%; Score 403.2; DB 1; Length 2067;
Best Local Similarity 92.8%; Pred. No. 3.1e-111;

Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;			
Qy	1	ATGAAACTTTTAAAGTCGGACATTCGAGCAATCGTAGTTTCTGGCAGTCTCTGGCT	60
Db	1193	ATGAAACTTTTAAAGTCGGACATTCGAGCAATCGTAGTTTCTGGCAGTCTCTGGCT	1252
Qy	61	GCGCGTATGATCAGCTGTTACCCGCTGTTTGTACCATGAAATGGCACATCATCCGCG	120
Db	1253	GCGCGTATGATCAGCTGTTACCCGCTGTTTGTACCATGAAATGGCACATCATCCGCG	1312
Qy	121	CCGAGCTCAACGCTTGAGCATTTATCAAGTACGGTTCCGCTAACGCTGCTCTGCA	180
Db	1313	CCGAGCTCAACGCTTGAGCATTTATCAAGTACGGTTCCGCTAACGCTGCTCTGCA	1372
Qy	181	AGCGATGCCCTTAAATCTGAAAACGACATTAACCGAGCGGTTATGCTAACCGCGCGCAT	240
Db	1373	AGCGATGCCCTTAAATCTGAAAACGACATTAACCGAGCGGTTATGCTAACCGCGCGCAT	1432
Qy	241	GTAGGCCAGGGTCGGATTAATAGTACTATGTAATGAACTGACTCAGAAATGTTTCAGAAATAT	300
Db	1433	GTAGGCCAGGGTCGGATTAATAGTACTATGTAATGAACTGACTCAGAAATGTTTCAGAAATAT	1492
Qy	301	GCCACCATCGAACCATGGAACGCTTAAAACTCCGATATTACTGTGCGCCAAATACGGCGGT	360
Db	1493	GCCACCATCGAACCATGGAACGCTTAAAACTCCGATATTACTGTGCGCCAAATACGGCGGT	1552
Qy	361	AATAACGGCGGCTGGTATATCAACCGCATTTGATTTCCAGGCTAATGCTGCTCAGGTT	420
Db	1553	AATAACGGCGGCTGGTATATCAACCGCATTTGATTTCCAGGCTAATGCTGCTCAGGTT	1612
Qy	421	GTTTTGGCAACACGCCACGGCTAACCAAGTATTAA	456
Db	1613	GTTTTGGCAACACGCCACGGCTAACCAAGTATTAA	1648
RESULT 3			
LOCUS	STAJ2301	5103 bp	DNA linear BCT 15-NOV-2000
DEFINITION	Salmonella typhimurium csgC, csgF, csgE, csgD, csgB, csgA, and csgG genes.		
ACCESSION	AJ002301		
VERSION	AJ002301.1 GI:2739232		
KEYWORDS	csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene; csgG gene.		
SOURCE	Salmonella typhimurium		
ORGANISM	Salmonella typhimurium		
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Salmonella.		
AUTHORS	1		
TITLE	Romling U., Bian Z., Hammar M., Sierralta M.D. and Normark S.		
JOURNAL	Curl fibers are highly conserved between Salmonella typhimurium and Escherichia coli with respect to operon structure and regulation		
MEDLINE	J. Bacteriol. 180 (3), 722-731 (1998)		
PUBMED	98117058		
REFERENCE	2 (bases 1 to 5103)		
AUTHORS	Romling, U.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-OCT-1997) Romling U., Department of Bacteriology, Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN		
FEATURES	location/Qualifiers		
source	1..5103		
	/organism="Salmonella typhimurium"		
	/mol_type="genomic DNA"		
	/strain="SR-11"		
	/db_xref="taxon:602"		
	<1..322		
CDS	/note="ORF179a"		
	/codon_start=2		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="CA05311.1"		
	/db_xref="GI:2739233"		
	/db_xref="GOA:O54290"		
	/db_xref="SWISS-PROT:O54290"		
	/translation="CTLSGMVMAVLVIIHGSALAPHLEIVSVYLTGIVAFIMCIGAKQ		
	LLTFVPGTFIGACATFAGCGDMRLVPLSLAGLIFGYAMKNSGIMLASRRRHSANT		
	AVTR"		
	complement (361..1194)		
	/gene="csgG"		
CDS	complement (361..1194)		
	/gene="csgG"		
	/function="lipoprotein"		
	/codon_start=1		
	/evidence=experimental		
	/transl_table=11		
	/product="CsgG protein"		
	/protein_id="CA05312.1"		
	/db_xref="GI:2739234"		
	/db_xref="GOA:O54291"		
	/db_xref="SWISS-PROT:O54291"		
	/translation="MPRLITIVAVLLSGCTAPKQAKPTLMPRAOSYKDLTHLPA		
	PTGKIFVSVNIGDETFQFYPASNSSTAVPSATMALVTALADSRWFTLEKQGLQ		
	NLNERKTIIRAQENGTVAMNRIPLQSLTANITMVGSIIGYSNVKSGGVARFYG		
	IGADTQYQLDQIAVNLRVNVSTGEIISVTSKTIISYEVQAVFFRIDYQRLIEG		
	IGYTSNBPVMLCLMSALETGVIFLINDGIDRLMDLQNKADRDNDILVYKRELAVPE		
	S"		
	complement (1221..1637)		
	/gene="csgF"		
CDS	complement (1221..1637)		
	/gene="csgF"		
	/codon_start=1		
	/transl_table=11		
	/product="CsgF protein"		
	/protein_id="CA05313.1"		
	/db_xref="GI:2739235"		
	/db_xref="SWISS-PROT:O54292"		
	/translation="MRVKAHVALLMLFSPITWAGNMTFOFNPNGNPNNGSFLNS		
	AAQNSYKDPAYNDPFIETPSPALDNFTQATQSOILGGLLTNINTGKGRVNTDFII		
	DIANDRQDLQNLVTRDKTGRTIEVSGLOIQTDF"		
	complement (1664..2059)		
	/gene="csgB"		
CDS	complement (1664..2059)		
	/gene="csgB"		
	/codon_start=1		
	/transl_table=11		
	/product="CsgB protein"		
	/protein_id="CA05314.1"		
	/db_xref="GI:2739236"		
	/db_xref="SWISS-PROT:O54293"		
	/translation="MKRYLTMVVAELLPATGNLANEVEVEVGLLTDHVSIGHE		
	FYRPFSPKSESYGNLTINERPARSGSWTTITVNDVIFQFTLFPKRPBKTIVF		
	ALADTERALNRQIDOTLLSTDLARDF"		
	complement (2064..2714)		
	/gene="csgD"		
CDS	complement (2064..2714)		
	/gene="csgD"		
	/function="transcriptional regulator"		
	/codon_start=1		
	/transl_table=11		
	/product="CsgD protein"		
	/protein_id="CA05315.1"		
	/db_xref="GI:2739237"		
	/db_xref="GOA:O54294"		
	/db_xref="SWISS-PROT:O54294"		
	/translation="MFNEVSSHGHTLLITKPSLOATVALLQHLKQSLATITKLIANTQ		
	RSLEVDISAGCIVLMDMMDADKLIHYQDNI SRKNNI KTLILNTPPDYRELIENP		
	HINGEVFATEDOEHVSGILGEGCYFOKLASYIITHSNRYNSTSGALLTHEE		
	KEILNKRLRIGASNNIEARSLFISENTVKTLYNLVFKKIAYGNRTQAVSWANDLNR"		
	complement (2869..2896)		
	/gene="csgDERG"		
	complement (2869..2896)		
	/gene="csgDERG"		
	complement (2891..2896)		
	/gene="csgDERG"		

```

gene
  3344..3376
  /gene="csGBA"
-35_signal
  3344..3349
  /gene="csGBA"
-10_signal
  3371..3376
  /gene="csGBA"
gene
  3469..3924
  /gene="csGBA"
  3469..3924
  /gene="csGBA"
  3469..3924
  /gene="csGBA"
  /function="nucleator"
  /codon_start=1
  /evidence=experimental
  /transl_table=11
  /product="CsgB protein"
  /protein_id="CA05316.1"
  /db_xref="GI:2739238"
  /db_xref="GOA:P55226"
  /db_xref="SWISS-PROT:P55226"
  /translation="MRKLIFMLITLIGAPGATATYDILARSEYFPAVELSKSFN
  QAAIQVGTDNARVROEGSKLIVIOEGNNNAKVDQNGNYPAYIFQTNANDA
  SISQSAVGNAAAIQKSGNKANITQYGTAVVOKOSHMAIRVQR"
  3469..3531
  /gene="csGB"
  3532..3921
  /gene="csGB"
  /product="CsgB protein"
  /function="nucleator"
  3966..4421
  /gene="csGA"
  3966..4421
  /gene="csGA"
  /function="fiber subunit"
  /codon_start=1
  /evidence=experimental
  /transl_table=11
  /product="CsgA protein"
  /protein_id="CA05317.1"
  /db_xref="GI:2739239"
  /db_xref="GOA:P55225"
  /db_xref="SWISS-PROT:P55225"
  /translation="MKLTKVAPFAIVSGSALAGVVPWGCGGNNHGGNSGSPST
  IDQNAKNSDITVQYGGNNALVNOTASDSVWVRQVGFNNATANQY"
  3966..4025
  /gene="csGA"
  4026..4418
  /gene="csGA"
  /product="CsgA protein"
  /function="fiber subunit"
  4483..4809
  /gene="csGC"
  4483..4809
  /gene="csGC"
  /codon_start=1
  /transl_table=11
  /product="CsgC protein"
  /protein_id="CA05318.1"
  /db_xref="GI:2739240"
  /db_xref="SWISS-PROT:P55227"
  /translation="MHTLLILAAISNOTITFTTQGGDIYVIVPVTLEPCVQVOLT
  SVRDVGCGSHQKQKTLSPANQPIELSRSLVINSRSDSVKLIIVTSDDQSILHSQ
  WPSAQ"
  4483..4533
  /gene="csGC"
  4534..4806
  /gene="csGC"
  /product="CsgC"
  4940..5103
  /gene="ORF103"
  4940..5103
  /gene="ORF103"
  /codon_start=1
sig_peptide
  mat_peptide
  gene
  CDS

```

```

ORIGIN
Query Match      86.7%; Score 395.2; DB 1; Length 5103;
Best Local Similarity 91.7%; Pred. No. 9.1e-109;
Matches 418; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

  1 ATGAACCTTTTAAAGTGGCAGCATTCGACGATGCTGTTTCGGCAGTCTGCTGCT 60
  3966 ATGAACCTTTTAAAGTGGCAGCATTCGACGATGCTGTTTCGGCAGTCTGCTGCT 60
  61 GCGCTTATGATCAGCTGGTTACCCGCTGTTTACCCATGAAATGGCAGATGATCCGGC 120
  4026 GCGCTGCTTCCACATAGGGGCGGCGGCGGATATCATTAACGGCGGCAATAGTCCGGC 4085
  121 CCGGACTCAACGTTGACATTTATGATGATGATGATGATGATGATGATGATGATGATG 180
  4086 CCGGATTCACGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4145
  181 AGCGATGCCGTTAAATCTGAAACGACCATTTACCGACGCGTTATGATGATGATGATG 240
  4146 AGCGATGCCGTTAAATCTGAAACGACCATTTACCGACGCGTTATGATGATGATGATG 4205
  241 GTAGCCACGAGGTTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 300
  4206 GTAGCCACGAGGTTGAGTAAATGATGATGATGATGATGATGATGATGATGATGATG 4265
  301 GCGACCATTCACGATGAAACGCTTAAACCTCGATATTAAGTTCGCAATACGCGCGT 360
  4266 GCGACCATTCACGATGAAACGCTTAAACCTCGATATTAAGTTCGCAATACGCGCGT 4325
  361 AATACCCCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 4385
  4326 AATACCCCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATG 4421
  4421 GGTTCGCAACGACCGCGCTTAAACCTCGATATTAAGTTCGCAATACGCGCGT 4485
  4386 GGTTCGCAACGACCGCGCTTAAACCTCGATATTAAGTTCGCAATACGCGCGT 4421

RESULT 4
AE008749 22411 bp DNA linear BCT 23-APR-2003
LOCUS      Salmonella typhimurium LT2, section 53 of 220 of the complete
DEFINITION
ACCESSION AE008749 AE006468
VERSION
KEYWORDS
SOURCE
ORGANISM
  Salmonella typhimurium LT2
  Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
  Enterobacteriaceae; Salmonella.
  1 (bases 1 to 22411)
  Mclelland M., Sanderson K.B., Spieth J., Clifton S.W.,
  Hou S., Layman D., Leonard S., Portolillo S., Ali J., Dante M., Du F.,
  Stenking T., Nhan M., Waterston R. and Wilson R.K.
  Complete genome sequence of Salmonella enterica serovar Typhimurium
  LT2
JOURNAL   Nature 413 (6858), 852-856 (2001)
MEDLINE   21534948
PUBMED    11677609
REFERENCE 2 (bases 1 to 22411)
AUTHORS   The Salmonella typhimurium Genome Sequencing Project
TITLE     Direct Submission
JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of

```

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest
Park Boulevard, St. Louis, MO 63108 USA
COMMENT Supported by NIH grant 5U 01 A143283

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs;
GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and
GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>,
and Pedro Romero and Peter Karp at EcoCyc;
<http://ecocyc.org/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites
were kindly provided by Heladia Salgado, Ulijo Collado-Vides and
ReguonDB;
http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one m13 subclone.

FEATURES

SOURCE

location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/db_xref="taxon:99287"

/note="LT2"

434..1308

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; RegulonDB:STMS1H001398"

454..1308

/gene="phoH"

/note="similar to E. coli Phob-dependent, ATP-binding pho
regulon component; may be helicase; induced by P
starvation (AAC74105.1); Blastp hit to AAC74105.1 (354
aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="Phob-dependent ATP-binding pho regulon
component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="WCAHNRVQENAGGVYVLQKLSNGROKAVIKARRAKRLRDS
RSHKQRESEVTSIAVQGVAEALIGMADSRDTSVYKARNEQAHYLAIDSKOIFAT
GEAGCGKTWISAAKAAALIHKDERIIVTRPVLADEDLGFLPDIAEKPAQYFRV
YDLIKRLGASFMQYCLRPEIKGYEIAFPAYMGRTFENAVVILDEAONTAAGMKMF
LTLLGENVTYVINGDITQCCLPRGVRSGLSDALERFEDEMGIVHFNKDDCYRSALC
QRTLHAYS"

complement(1414..2303)

/gene="STM1127"

complement(1414..2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein
(AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcriptional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="WEPQPRPKKGLLDLGLMGOKSLTPASORIAQYITAFPROVTO
SSIAIDSRTOAGATVIRCRITGVGFDFKDLAIELATTESDDSSPLDAEVS
SDHAHAIGLKQNTISNVLETNLDMQOVLGVADLRCHSVYIFGVSSGTTALD
MKHLMRIGLRGDAVSNHFMVQATLLKAGDAVMGVSHSGTSPETVHSIRLARQAGA

RBS

TTVAITNIGSPICEADFCILINGNRQMLQGDISTKRAQLFVFDLLYTLVQSSPE
OARBSKIRTNALDMTK" complement(2298..2303)

/gene="STM1127"

/note="putative RBS for STM1127; RegulonDB:STMS1H001399"

complement(2580..4088)

/gene="STM1128"

complement(2580..4076)

/note="similar to E. coli putative cotransporter
(AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%

identity in aa 7 - 478"

/codon_start=1

/transl_table=1

/product="putative sodium/glucose cotransporter"

/protein_id="AAL20059.1"

/db_xref="GI:16419644"

/translation="MITHSFGIVYVLYLFGYLLAMLVGVYFSRRROKADDPFGGGR
VPGWAGVSPATLTSITFMSIPAKFTSDMTFIIQOVLAIPLVYFYIPFRK
LKTSAYEYIEARFDVRCRLPASMFLPHIGRAITFTLVLRPPIADPILVLA
LISVWCITVYMWGGIGBVITVDVIGLLSGSALIFIVICLKVOG61DEIFVTQOL
DKFPPATQFMSWSTFVPMIGLFLFANTIOQFASQDVQVRYVTSIDETKTLT
NAKLAVIYVPEFPAIGSALFVYQOHPQLPAGNTGILPLFVTEMVPAIGLIIA
ALPAAQSSISSINSISGFSNDIYORLSHKRTPENRMLAKLVILVIGISSAS
VLMVADESRTPAFNSILGMPMTGLFMTGIFFRANAGSAVLG11ISVITVLA
RYATDLNFPFYGVIGSLSVIGVTFAPLFAFAPLTLDEKPEKVTL"

complement(4083..4088)

/gene="STM1128"

/note="putative RBS for STM1128; RegulonDB:STMS1H001400"

complement(4413..5099)

/gene="STM1129"

complement(4413..5093)

/gene="STM1129"

/note="similar to E. coli putative enzyme (AAC76255.1);
Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -
226"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20060.1"

/db_xref="GI:16419645"

/translation="MSILARLEOSVHENGGLIVSCOPVSPMDKPEIYVAAQAAAS
AGAVAVRIGELIENRTVRLSVPIIGIRDLTGSVRITPVYQDVALAQAADII
APDAPRSRRVDLSLIRIRINGLLAMACSTYNEISCHQKIEITGLTSYTCQ
ITPEPDLAVTQDSHAGCVIABGRINTPLAANAIEHGAIVATVGSALTIRSHICQ
WFSHAVER"

complement(5094..5099)

/gene="STM1129"

/note="putative RBS for STM1129; RegulonDB:STMS1H001401"

5599..6759

/gene="STM1130"

5599..6759

/gene="STM1130"

/note="similar to E. coli orf, hypothetical protein
(AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

identity in aa 40 - 403"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20061.1"

/db_xref="GI:16419646"

/translation="MGMOKNPKKMTIALLCSVAITTSQVATTLDPIDPPLKNGC
AIDNIGYVGLDTAGISWYKIDLKQKHQWERIKSPGARGOSVAFNLDELVYVG
GVGRKNSBPLQVYSDVYKXSPVKNWQKDTISPVGLTRKLNELTVLLITGVN
EHIIFDKYFIDIAAADADESEKNTVINYPAKPAKADYFENKIVFIYNAKENTYKAGSL
PGAGTASSSSVMGNFIMLINGELKGLRDIYVYRAMNDKLTMLKNSQLPSPGEO
MOEGLAGFSGYSGHVLVGGGALPGKAGKNTYNGKYSRGIINKKRDVYGLINSH
MOYMGKMKQPLIGVSVSYSDVPLIGENAKKRPVSYSFTMRGNLLIK"

6793..7497

/gene="STM1131"

6793..6798

/gene="STM1131"

/note="putative RBS for STM1131; RegulonDB:STMS1H001402"

gene

RBS

CDS

gene

RBS

CDS

gene /note="PS00142 Neutral zinc metalloproteinases, zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:O82759"
/db_xref="SWISS-PROT:O82759"
/translation="MYPRKALFQIDPRAHEFTFOQLRITGTPLEALVROKVP
KPTCMGLTFNPLGLAGLDKDEICDALAMGFSLEIGTVPRPQNDKPLFR
LVDAEGLINRGNFNNLVNENKKAHPGILGINRKNKTPVENKQDYLIME
KVAVAGYIAINISSPNTPLRTLOYGADLDLITAIKNKNDLOVHHKYPVAVKI
APDICEELIOVADSLRHNDVIAATTTIDRSIYQMKRCCOTGSLGSRPLDKST
EIRRLSELKQGLPIIGVGIDSVIAREKIAAGATLVQIYSGPIFKGPLKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHODHase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBW_ECOLI; Faeta hit to YCBW_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTRMBL:O82758"
/translation="IMRIKIPDDNRWYDEEDHMMDLNGLFRSFRKMLTPDA
FCPTGCVDDALYFSEKCRDEPLTEQRAELVNLNVAIRYLKPPMPSMFAVH
GEMWTPGIDGASVWLSDTAEQVNLVVEPEENALCLLAPGVVINGRTMQLDALK
IMNRLKPOVHCHSFSLQAV"
gene complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:O82757"
/db_xref="SPTRMBL:O82757"
/translation="MATLSRLFIHPVSMGIGLTHALDISGLAFDRIMVTSDDT
FITARQPPWRTFSPPLHDLITAPDGSALVRFDFPQDPTVGMNHFAFARA
PTAINOMLSGFSFSDVOLCWGPOLTRRKHNAVPLGADGYLLTNESSLDGO
RCPAGVMOEORPMLVSVGAAMEDSMKVLRIGDVI EDVVKPSCRICFTTSPSEKO
KHSGEPLATQAFTADNDNDUNFGONLIARNSGVIRVGDVEVILATAKAGATT
LDBSVTPKRPAGDSVTTIDWQQTFCGNNOVLLELSENGIRIYSCRAGICGCRIR
LLEGEVPLKKSAMGSDGTILSCSVKTAIRLEN"
misc_feature complement(4737..4955)
/gene="STY1081"

gene /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein H10116 H10116 SW:YCBY_HAEIN (P44524; P43945) Faeta scores: E(): 0, 60.3% id in 710 aa Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:O82756"
/db_xref="SPTRMBL:O82756"
/translation="MNSLPASTAGKEELIKIEKLGAVGCVVGVHROGTRLI
YQSLMRSRLSRILPMGECKVSDLDLYEVQALNTEIFNPAITRAVHFSGLNDTI
RNSQYGAMKVDALVDAFTRKNLPRPVNDRESPLRLNWLKNTASIALDLSGLH
LRGYRDRITGLAPIKETLAAIYVMSGQPGTPLDPMCGSGTLLIEAMWATDRAPIGL
HRCWMSFGMAQHDDEAIWQEKVAPQTRAKGLAEYSHSFYGSDSARVIRASNNAR
RAGIGELITFEVVDVAQNSNPLPKGPVTSNPGYERLSDPALIAHSLIGRTYK
NORGKMLSLFSAVSDLSGLADQAFKANGPLDVCOKNTHIATITADSKATYA
EDVANRLKDLKLEKVARQSGIECYRLDYDDEEVNADRYDGMATVOTAPKTV
DAQVAKRQLPDIIATLSVAGIIPNNKILKTRERQKNDQYKMSGEFLVESEYNA
RLWNLITDYDGTGLFDHRIARMLGMSKQPLANFSYGSASVHAGLGAGANTTT
VDMSTRYLEMAERMLRLNGRAHRLIQDQCLMLBEANQFPLIFIDPTPFNSKR
MESEFDVORHVALMKDLKRLKRGKGTIMSNNGRPRMDLEGLAELGTAEITQKT
LSPDFANROIHNKMLRAA"
misc_feature 5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Faeta hit to YJYU_ECOLI (554 aa), 34% identity in 524 aa overlap
Query Match 86.3%; Score 393.6; DB 1; Length 254050;
Best Local Similarity 91.4%; Pred. No. 3.9e-108;
Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGAGCATTCGACGATCGATGTTTTCGACGCTCGGCT 60
Db 88906 ATGAACTTTTAAAGTGGAGCATTCGACGATCGATGTTTTCGACGCTCGGCT 88965
QY 61 GCGCTCTATGATCAGCTGTATCCCGTGTGTTTACCATGCAATGACATGATCCGGC 120
Db 88966 GCGCTCTTCCACATATGGGCGGGCGGTATCATTAACGGCGGCAATAGTTCCGGG 89025
QY 121 CCGGACTCAACGTTGAGCATTTTATCATGATGCGTTCCGCTAACGTTGGCTTCTCTGCA 180
Db 89026 CCGGATTCACAGTTGACATTTTATCATGATGCGTTCCGCTAACGTTGGCTTCTCTCAA 89085
QY 181 AGCGATCCCGTAAATGTAAGACATTTACCAAGACGGGTTATGTTAAGCGGCCGAT 240
Db 89086 AGCGATCCCGTAAATGTAAGACATTTACCAAGACGGGTTATGTTAAGCGGCCGAT 89145
QY 241 GTAGCCAGGAGGTGGATTAAGTACTATTGAACCTGACAGATGTTTCAGAAATAT 300
Db 89146 GTAGCCAGGAGGTGGATTAAGTACTATTGAACCTGACAGATGTTTCAGAAATAT 89205

REFERENCE	AUTHORS	TITLE	JOURNAL	PUBLISHED	REFERENCE	AUTHORS	TITLE	JOURNAL	PUBLISHED			
1	Deng, W., Liou, S.R., Plunkett III, G., Mayhew, G.F., Rose, D.J., Burland, V., Kodoyanni, V., Schwartz, D.C. and Blattner, F.R.	Comparative Genomics of <i>Salmonella enterica</i> Serovar Typhi Strains T2 and CT18	J. Bacteriol.	185 (7), 2330-2337 (2003)	2	Deng, W., Liou, S.-R., Plunkett, G., III, Mayhew, G.F., Rose, D.J., Burland, V., Kodoyanni, V., Schwartz, D.C. and Blattner, F.R.	Direct Submission	Submitted (25-SEP-2002)	3	Wisconsin - Madison, 445 Henry Mall, Madison, WI 53706, USA	Location/Qualifiers	1. .301983

CDS

complement(164..622)
/gene="tnpA"
/locus_tag="t1735"
note="corresponds to STY1225 from Accession AL513382;
Salmonella typhi CT18"
/codon_start=1
/transl_table=1
/product="transposase for insertion sequence element
IS200"
/protein_id="AA069359.1"
/db_xref="GI:29137798"
/translation="MGDEKHSILWHTRNCKYHIVAPKYRQAFGEKKRAVGSII
CEMGNVALIEAECACAHIMHLLEIPKQSVSFMVGIKSSIMLYEONGDLKPF
REMFCEKRYVDYDTGKNTAKTIDYIHQLEEDKMGESQSLPYSPPTGRK"
/gene="f19L"
/locus_tag="t1736"
complement(844..1797)
/gene="f19L"
/locus_tag="t1736"
/note="corresponds to STY1223 from Accession AL513382;

Salmonella typhi CT18"
/transf_table=11
/codon_start=1
/product="Flagellar hook-associated protein 3"
/protein_id="AA069360.1"
/db_xref="GI:29137799"
/translation="MRISTOMVYFQNMSCGTTNQDAWMKLGQMGSTGRKRVNPPD
ASQAQAVLPLQAOAONSQYVALRATRTQKVSLEESVLSQVTTIQTAKRKIYVAGNGT
SDDRRASLTLDLQIGRIDQMLNLANSIDGNKRYIPACYKTEALPFOAATGAGKEKES
TQGVDAFVWVGHGTGAQATPSTISLVNVPDSDSEKALFMLDLPTALAAIKTVEEN
DYVEKKAALAIIDKTKRGLKSLNNVLVRALELTQISELTSTDSGSRALGQQLQNS
NLVDYDMSVSVISYMQOAPLQASVYKFTFDMQMSLFLQNR"
/gene="flgK"
/complement(1812..3473)
/locus_tag="ct1737"
/locus_tag="ct1737"
complement(1812..3473)
CDS

gene
VQDSVSGNLTBEYGNQRYQOQYYLNAQVLOTATNALFDALNTRIS" complement (3538. .4488)
/gene="flgJ"
/locus tag="tl738"
/gene="flgJ"
complement (3538. .4488)
/gene="flgJ"
/locus tag="tl738"
/note="Corresponds to STY1221 from Accession AL513382:
Salmonella typhi CT18"
/codon start=1
/transl_table=11
/product="flagellar protein FlgJ"
/protein_id="AA069362.1"
/db_xref="GI:29137802"
/translation="MIDGGKLLASANDAQSLNELKAKAGODPAANIRPVARQYEGMF
VOMMLSMELPAKXGDLSPSDQRLVYTSMDQIQAOMAGGGLAEMWKMWTGQ
TMDPDAPOVPLKSLFETVNSYONQALTVLVRALIKTDSDDALSGDSMDPLRLRS
LPLRLASBQSGVHHLLIAQNALSGMGQCFQILRENGEFSNVVPCVATKSGKPYTE
ITTTRENSBAKRVAKARVYSSYLEALSPYALLTRNPRYAATTAITAEQGVALQI
NAGIADLPYAKVLAAMQOLKAMSEKVSKTSAINDLNF"
complement (4488. .5591)
/gene="flgI"
/locus tag="tl739"
complement (4488. .5591)
/gene="flgI"
/locus tag="tl739"
/note="Corresponds to STY1220 from Accession AL513382:
Salmonella typhi CT18"
/codon start=1
/transl_table=11
/product="flagellar P-ring protein precursor"
/protein_id="AA069363.1"
/db_xref="GI:29137802"
/translation="MYVFKALAGITVALVATLAHERIRIDTISVQGRNSLIGYIV
VGLGSDQDTQPTPTTQTLNNMLSQIGITVPTGNQLKNVAWVTSYPPAPROC
OLTIIVVSSMKNKSLRGSTLLMTPLKGVSSOYALAQGNIIVGASAGSGSSQVN
OLNGCRITNGALITREPLPTQFAGNTINLQINBEDFTMAQOILTDINARGGISTAL
DARTVQRPSNGSSGQFVLADIQMEENVATPDPAKVINSTRGSSVWNRREVTILDSCA
VAQGNLSVTVNRQLNVNQPTPTGGGCGIVVTPTQIIDLQSGGSLQSVRSSANLNSV

gene
RAINAATGATPMDLMSILQSMOSAGCLRAKEIT
complement (5597..6295)
/gene="f1gH"
/locus_tag="t1740"
complement (5597..6295)
/gene="f1gH"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQKYLALHPYPMALMTVLTGCAWIPAKPLVQCATTTAQTPIGPY
PVANGSIFOSAPINTYGTQPLFEDRRPNITDITLTIVLQENVASKSSAASRDGKT
SFGDTPRYLQGLFGNSRADMEASGSGSFGKGAASNTFSGTLTIVQVLANGN
LHVVEKQIALINQCTEFIRFSGVNPRTISGNSVSPSTQVDARIEYVGNQYINDAQ
MGMORFPLNLSPM"
complement (6350..7132)
/gene="f1gG"
/locus_tag="t1741"
complement (6350..7132)
/gene="f1gG"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal rod protein f1gG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISSLRITAKTGGLDAQCTNMVIANNLAVSTNGFKRORAVEDL
LYQTRKPGASSTQTLPSGLQIGTVRPATERLHSQGLSQTNNSKDAIKGQGR
FOVWLPGSTAYSTRDGSPQVQNGQLYAGAFQVQPAITTPANALSTIGRQVSVT
QQCGAAYVQVQLNLTTFMNDTGLSISGENIYETSSGAPNESTPGINAGLHYGY
VETSNVVAELVNMIOVQRAYEINSAVSTTDQMLQKLTQL"
complement (7146..7901)
/gene="f1gF"
/locus_tag="t1742"
complement (7146..7901)
/gene="f1gF"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein f1gF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHAIITAMGASQTLNQAVTNSNLANASTPGFRADNLRAV
PVDLSLATRTLVASTPGADMTQGLDYTSRPDLVALQODGMVVOAADAEGYTN
GNLVQPTQGLTIQGHPIVIGEGPITVPEGEIRIADGTSALNPGSPPTVAIVGR
LKVYKAGNEVQRSDGFLRLTAQAQRAGLADSIIRMSVILSGSNVKEVEMAT
DMINARFEMQKRVITSVDENGRANQLMSMS"
complement (7922..9133)
/gene="f1gE"
/locus_tag="t1743"
complement (7922..9133)
/gene="f1gE"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein f1gE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFSQAYSGMAATNIDVGNINANSATYGFKSPADMA
GSKVGLGVKAGITQDFTDGTITNGRLDVALISONGPRLVDSNQSIVFYSRNQFTL

gene
86.3%; Score 393.6; DB 1; Length 301983;
Best Local Similarity 91.4%; Pred. No. 4e-108;

Matches 417; Conservative 0; Mismatches 39; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGCAGCATTTGCGACCAATCGTAGTTTCTGGCAGTCTTGCT 60
DB 37310 ATGAACCTTTAAAGTGGCAGCATTTGCGACCAATCGTAGTTTCTGGCAGTCTTGCT 37251
QY 61 GGGCTTATGATCGAGCTGTTACCGGTTGTTTACCATGAAATGCGCATGCAATCCGGC 120
DB 37250 GGGCTTATGATCGAGCTGTTACCGGTTGTTTACCATGAAATGCGCATGCAATCCGGC 37191
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGGTTCGGTAAAGCTGCGCTTCTGCGAA 180
DB 37190 CCGACTCAACGTTGAGCATTTATCAGTACGGTTCGGTAAAGCTGCGCTTCTGCGAA 37131
QY 181 AGCGATGCGCGTAAATCTGAAAGCAGCATTAACCAAGCGGTTATGTTAAGCGCGCGAT 240
DB 37130 AGCGATGCGCGTAAATCTGAAAGCAGCATTAACCAAGCGGTTATGTTAAGCGCGCGAT 37071
QY 241 GTAGCGCAGGCTGGGATTAATAGTACTATTGAACTGACCTGCAAGTGGTTTCAGAAATTAAT 300
DB 37070 GTAGCGCAGGCTGGGATTAATAGTACTATTGAACTGACCTGCAAGTGGTTTCAGAAATTAAT 37011
QY 301 GCCACATCGACGACGAGTGAACGCTAAAACTCCGATATTACTGTCGGCCATATAGCGCGT 360
DB 37010 GCCACATCGACGACGAGTGAACGCTAAAACTCCGATATTACTGTCGGCCATATAGCGCGT 36951
QY 361 AATAAGCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGTT 420
DB 36950 AATAAGCGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGTCAGTT 36891
QY 421 GGTTCGCAACAGCGCAGCGCTTAACCATTA 456
DB 36890 GGTTCGCAACAGCGCAGCGCTTAACCATTA 36855

RESULT 7
STAGEBA
LOCUS STRAGEBA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AU000514
VERSION AU000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE Salmonella typhimurium
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE
AUTHORS Sukupolvi,S.S.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllynkatu, 20520, FINLAND
FEATURES
source
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/locus_tag="t1743"
/codon_start=1
/transl_table=11
/product="agfB protein"
/protein_id="CA04150.1"

```

/db_xref="GI:2275120"
/translation="MKMKLLFMMLTILCEPIATATNYDLARSEYNFAVITLSSKSPNO
AATIGVGTDNARVRHSGSLSVISQERGNARAKVDAGNYPAYITQGTNADA
SISQAYGNSAISAIIOKSGNKNANTYGTCTKAVVVOXKSHMAIQANTYGTQ
KTAVVVOXKSHMAIRVTR"
593..1048
/gene="agfa"
593..1048
/gene="agfa"
/codon_start=1
/translation="agfa protein"
/protein_id="CA04151.1"
/db_xref="GI:2275121"
/translation="MKLLKVAAPFAIVSASAVAGVVPOMGGGNNHGGSSGPDST
LSIYQYSANALALOSDARKSETTITGSGVGNADVGAGADNSTITELTONFRNNAT
IDWNAKNSDITVGOYGNNAALVNOTASDSVWVRQVGFNNAPANQYN"

```

Query Match 85.0%; Score 387.8; DB 1; Length 1048;
 Best Local Similarity 90.8%; Pred. No. 1.4e-106;
 Matches 413; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACATCGTAGTTCTGGCAGTCTGGCT 60
DB 593 ATGAACTTTTAAAGTGGCAGCATTCGACATCGTAGTTCTGGCAGTCTGGCT 652
QY 61 GGGCTATGATGAGTGTGTTACCGGTGTTTACCCAGTAATGCAATGATCCGGC 120
DB 653 GGGCTGTTTCAAAATGGGCGCGGCTTAATCAATACGGGCGGCAATGTTCCGGC 712
QY 121 CCGGACTCAACGTTTGAATTCATGATCGGTTCCGTTAAACGCTGCGCTTCTCGAA 180
DB 713 CCGGATTTCCACGTTGACATTTATCAATGATCGGTTCCGTTAAACGCTGCGCTTCTCGAA 772
QY 181 AGCGATGCCCGTAAATCTGAACGACCATTAACCCAGACGGTTATGTTAACGGCCGAT 240
DB 773 AGCGATGCCCGTAAATCTGAACGACCATTAACCCAGACGGTTATGTTAACGGCCGAT 832
QY 241 GTAGCCGAGGTTGCGGATTAATGTAATTAAGTGAATGTTGTTTCAAGAAATAT 300
DB 833 GTAGCCGAGGTTGCGGATTAATGTAATTAAGTGAATGTTGTTTCAAGAAATAT 892
QY 301 GCCACATCGACCACTGGAACGCTAAATCTCCGATTTTACTGTCGCGCAATACGGCGGT 360
DB 893 GCCACATCGACCACTGGAACGCTAAATCTCCGATTTTACTGTCGTCATACGGCGGT 952
QY 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGGTGGCTGAGTT 420
DB 953 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGGTGGCTGAGTT 1012
QY 421 GGTTCGCAACCAACGCGCGCTAACCACTATTA 455
DB 1013 GGTTCGCAACCAACGCGCGCTAACCACTATTA 1047

```

RESULT 8
 LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
 DEFINITION Sequence 56 from patent US 5635617.
 ACCESSION 144908
 VERSION 144908.1 GI:2469621
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 361)
 AUTHORS Doran, J. L., Kay, W. W., Collinson, S. Karen, and Clouthier, S. C.
 TITLE Methods and compositions comprising the agfa gene for detection of
 JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

FEATURES
 source location/Qualifiers
 1..361
 /organism="unknown"
 /mol_type="unassigned DNA"

Query Match 62.3%; Score 284.2; DB 6; Length 361;
 Best Local Similarity 99.0%; Pred. No. 4.3e-75;
 Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY 110 ATGCATCCGCGCCGACTCAACGTTTATAGTACGTTCCGTTAAACGTCGCGC 169
DB 47 ATATTCCTCCGCCCGACTCAACGTTTATAGTACGTTCCGTTAAACGTCGCGC 106
QY 170 TTGCTCTGCAACGCGTCCGTTAAATCTGAACGACCATTAACCGAGCGGTTATGTA 229
DB 107 TTGCTCTGCAACGCGTCCGTTAAATCTGAACGACCATTAACCGAGCGGTTATGTA 166
QY 230 ACGCGCGCATGTAAGGCGAGGTTGCGGATTAATGTAATTAATGTAATGTAATGTA 289
DB 167 ACGCGCGCATGTAAGGCGAGGTTGCGGATTAATGTAATTAATGTAATGTAATGTA 226
QY 290 TCAGAAATATATCCACCATTCACAGTGAACGCTAAATCCGATTAATCTGTCGCGC 349
DB 227 TCAGAAATATATCCACCATTCACAGTGAACGCTAAATCCGATTAATCTGTCGCGC 286
QY 350 AATACGCGCGTAAATACCGCGCTGTTAATCAAGCCGATCTGATTC 398
DB 287 AATACGCGCGTAAATACCGCGCTGTTAATCAAGCCGATCTGATTC 335

```

RESULT 9
 LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter sp. Fec2 csgB gene, csgA gene and csgD gene.
 ACCESSION AJ515700.1 GI:31790491
 VERSION AJ515700.1 GI:31790491
 KEYWORDS csgA gene; csgB gene; csgD gene; curlin-csgA protein; nucleation
 component of curlin monomers; regulatory protein.
 SOURCE Citrobacter sp. Fec2
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 REFERENCE 1
 Zogaj, X., Bokranz, W., Nintz, M., and Romling, U.
 Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 Infect. Immun. 72 (7), 4151-4158 (2003)
 JOURNAL 2 (bases 1 to 2889)
 AUTHORS Romling, U.
 TITLE Direct Submission
 JOURNAL Submitted (11-NOV-2002) Romling U., Microbiology and Tumoriology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source location/Qualifiers
 1..2889
 /organism="Citrobacter sp. Fec2"
 /mol_type="genomic DNA"
 /isolate="Fec2"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:213763"
 /country="Germany"
 /complement(212..862)
 /gene="csgD"
 /complement(212..862)
 /gene="csgB"
 /function="necessary for transcription of the csgBA
 operon"
 /codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="CAD56670.1"
 /db_xref="GI:31790492"

	gene	/translation="MFNFESHGHGTLTLLTRKPSLOATPLLQHKOSALSGKLHNLN RSLDIDSSCYLVDMMEADKKLHYKNODNSRKNNLTILITLPDYPRDIEMENI HINGFVYADDERVNGIQGLRGECFPSQLASYLITHSGNFRNPSESSALLTHRE KEILNKRLIGASNIEBIARSLFISENTVHTYNLFKIAIVNRTOAVSWANDINLR"
CDS	gene	1623.. 2078 /gene="cs9B" 1623.. 2078 /gene="cs9B" /function="nucleator" /transl_table=1 /codon_start=1 /product="nucleation component of curlin monomers" /protein_id="CAD56571.1" /db_xref="GI:31790493" /translation="MKNLPLFMMLTVLGAPGISATSYDLAHSFYNAVNELSKSPFN QAAILGOVGTTNNSAKMRQESKLSVSVEQGSNBAKDQSGAYVFAYIAOGRSHNDSD SISSNYGNTAMTIIOKSGSNKANIIYOYTGTUVAVVYQRQSOMAIRVIOR" 2119.. 2571 /gene="cs9A" 2119.. 2571 /gene="cs9A" /function="curli subunit" /codon_start=1 /transl_table=1 /product="curlin-csgA protein" /protein_id="CAD5672.1" /db_xref="GI:31790494" /translation="MKLIQVAFAPAIVVSGSLAGSVPMQGCGGGGSSGSESTSI SIVOSGVNMAILOSDARKSDPTIHQGFPGADVQSGSNSITLDLTONFPKNNATT DOMXKNSDIRVSYQYGHNAALNVNOTABDSVLHVQVDFGNMATTANQ"
CDS	gene	
ORIGIN		
Query Match	61.1%; Score 278.4; DB 1; Length 2889;	
Beet Local Similarity	77.2%; Pred. No. 3e-73;	
Matches 352; Conservative	0; Mismatches 101; Indels 3; Gaps 1,	
Oy	1 ATGAACCTTTAAAGTGCGACCATTCGCCCAATCGTAATTTTCGCGAGTCTTGECT	60
Db	2119 ATGAAACTTTTAAACAATGGCAGCATTTTCACCAATCGTGTTTTCTGCGAGCTTGECT	2178
Oy	61 GGCGTCTATGATCAGCTGGTACCCTGTTGTTGCATCCATGAAATGGCACATGCATCCGGC	120
Db	2179 GG---TTCTGTTCGCCAATGGGGCGGGCGGTGCGGGCGGGGAGAAGCTTCGGC	2235
Oy	121 CCGGACTCAACGTTGAGCATTTATCAGTAGCGTTTCCGCTAACCGCTGCGCTTGTCTGCAA	180
Db	2236 CCGGAATGCAACCTGAGCATTTATCAGTCAGAGATCAATAAGCGCGCTTGCCTGCAA	2295
Oy	181 AGCGATGCCGCTAAATCTGAAAACGACCATTCACCCAGAGCGGTTATGTAACGGCGCCGAT	240
Db	2296 AGCAGCGCTCTTAATCTGATACGACCAATTCATCAGTAATGCTTGTGTAAACGGCGCAGAC	2355
Oy	241 GTAGGCGCGGGTGGCGGATATAGTACTTGAACGATCGAANTGTTTCAGAAATAT	300
Db	2356 GTGGGCCAGGGCTCAGATPACGACCATCGATCTGACTCAAAACGGCTTCAAAAAAC	2415
Oy	301 GCCACCATCGACCAETGAGACGCTAAAACTCCGATATTACTGTGCGGCAATACGGCGGT	360
Db	2416 GCACACATCGATCAGTGAACGGCAAAATTCGGAATTAATCTGTAGACAGATATGTGGA	2475
Oy	361 AATAACCGCGCGCTGTATTATTCAGACCCGATCTGATTCAGCGTAATGTGTCGTACAGTT	420
Db	2476 CATTAACCGCGCATCTGTGTGACACAGCTGCGTCCGATTCAGACGTTCTGTGATCAGGTT	2535
Oy	421 GATTTTGGCAACAAGCAGCGGCTTAACAAGATTAA	456
Db	2536 GATTTTGGCAACAAGCAGCGGCTTAACAAGATTAA	2571
RESULT 10		
CFRS15701	2920 bp DNA linear BCT 24-JUN-2003	
LOCUS		
DEFINITION	Citrobacter freundii csgB gene, cs9A gene and cs9D gene.	

Accession	Version	Keywords	Source	Organism	Reference	Authors	Title	Journal	Reference	Journal	Authors	Title	Journal	Features	Source																																						
AJ515701	1	GI:31790495	csga gene; csGB gene; csGD gene; curlin-csgA protein; nucleation component of curlin monomers; regulatory protein.	Citrobacter freundii	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.	Zogaj, X., Bokranz, W., Nimtz, M., and Romling, U.	Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)	2 (bases 1 to 2920)	Direct Submission	Submitted (12-NOV-2002)	Romling, U., Microbiology and Tumourbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN	Location/Qualifiers	1. .2920	/organism="Citrobacter freundii"	/mol_type="genomic DNA"	/isolate="Rec4"	/isolation_source="faeces"	/specific_host="Homo sapiens"	/db_xref="taxon:546"	/country="Germany"	complement(211. .861)	/gene="csGD"	complement(211. .861)	/gene="csGD"	/function="necessary for transcription of the csGA operon"	/codon_start=1	/transl_table=1	/product="putative regulatory protein"	/protein_id="CAD56673.1"	/db_xref="GI:31790496"	/translation="MFNEVSHGHHTLLITKPSLQATPALQHLKOSIALTGKLTNLICRSLLDSSICVLDDMMADKILSVISGEGSNRAKTDOTGSYNFAVIDOTSSNDHINCFATPEDEORVVGLOGVLRGEYFSGKSLASYLITSGNRYNSTESALLTHREKEILANKRIGASNEIRASLPISENVKTHLYNLFKLIAYNRIGQAVSMANDNLR"	1624. .2082	/gene="csGB"	1624. .2082	/gene="csGB"	/function="nucleator"	/codon_start=1	/transl_table=1	/product="nucleation component of curlin monomers"	/protein_id="CAD56674.1"	/db_xref="GI:31790497"	/translation="MKRYKLLPMFPMGLGVGCIAAATTSYDLANSEYPAVNELSKSSNOAIIITQVGTANSANTROGSKILSVISGEGSNRAKTDOTGSYNFAVIDOTSSNDASIKQSGYNTVAIVIQKSGNKANITQYGTQTAIVVOROSQMAIRVTR"	2123. .2572	/gene="csGA"	2123. .2572	/gene="csGA"	/function="curli subunit"	/codon_start=1	/transl_table=1	/product="curlin-csgA protein"	/protein_id="CAD56675.1"	/db_xref="GI:31790498"	/translation="MKLIKTAAPRAIIVSSGSLAGVVPQWGNHGGSGSYGPPSSLSITVQGSNNSANTROGSKILSVISGEGSNRAKTDOTGSYNFAVIDOTSSNDIQMAKNADINLTQFGERNGLVNOTGADSNVLLIQVGFNNATANQH"

QY	1	ATGAAACTTTTAAAGTGGACGACATTCGACCAATCGTAGTTCTGGAGTGTGCT	60
Db	2123	ATGAAACTTTTAAAGTGGACGACATTCGACCAATCGTAGTTCTGGAGTGTGCT	60
QY	61	GCGCTATGATTCAGCTGGTATCCCGTGTGTTACCATGAAATGCGACATCGCCG	120
Db	2183	GCTGTGTTCCGCAATGG-----GGGCAATCATCATGTGTGGCAGTAAATTATGCG	120
QY	121	CCGACATCAACGTTGAGCACTTTATTCAGTACGCTTCGGCTAAGCGTCCGCTGCTGCA	2236
Db	2237	CCGACATCTTCACGAGTATCTTCAATTTGGTCAAACTCTGCAAAATGCTCTGCA	180
QY	181	AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGTTATGTATACGGCGCGAT	2296
Db	2297	AGTATGCGCGTAAATCATGATGTCACCATCAACAACGGCGGTGGCAACGAGCTGTT	240
QY	241	GTAGGCGAGGTCGCGATTAATAGTACTATTGAATGACTCGAGATGGTTTCAGAAATAT	2356
Db	2357	GTTGGCGAAGGTCGTATGACAGTACCTTTAGCGTGAACAGACCGGCTTCCAGAACGT	300
QY	301	GCCACCATGACACAGTGGACGCTPAAAACTCCGATATTACTGTGGCCATACGGCGGT	2416
Db	2417	GCCACTATCGATCAGTGGAAATGCAAAAAATGCTATTTATGCGTGACCCAGTTCCGTGGC	360
QY	361	AATAACGCGCGCTGGTTAATCAACCGCATCTGATTCAGCGCTAATGGTGGCTCAGGTT	2476
Db	2477	CGCAACGGTGGTGGTTAATCAACACGCGCTCGACTCCAAATGCTGATTACGACAGGTT	420
QY	421	GTTTTGGCAACAACGCCACGGCTAACCACTATTAA	456
Db	2537	GGCTTTGGTAAACAAGCTACAGCTAACCAACAATTAA	2536

FEATURES	source
LOCUS	AF275733
DEFINITION	AF275733 1711 bp DNA linear BCT 15-MAY-2000
ACCESSION	AF275733
VERSION	AF275733.1
KEYWORDS	GI:14039399
SOURCE	Escherichia coli
ORGANISM	Escherichia coli
REFERENCE	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.
AUTHORS	1 (bases 1 to 1711)
TITLE	Unlich,G.A., Keen,J.E. and Elder,R.O.
JOURNAL	Mutations in the csgr promoter associated with variations in curli expression in certain strains of Escherichia coli O157:H7
REFERENCE	Appl. Environ. Microbiol. 67 (5), 2367-2370 (2001)
JOURNAL	MEJLJNH
AUTHORS	PUBMED
TITLE	11319125
JOURNAL	2 (bases 1 to 1711)
REFERENCE	Unlich,G.A., Keen,J.E. and Elder,R.O.
AUTHORS	Direct Submission
TITLE	Submitted (06-JUN-2000) USDA, ARS, Roman L. Hruska U.S. Meat Animal Research Center, State Spur 18D, Clay Center, NE 68933, USA
LOCUS	1. 1711

/db_xref="GI:14039400"
 /translation="MKKKLLFMWLTILGAPGIAAAGYDLANSEYFANVELSKSFFN
 QAAITIGQAGTNNASQLRQSGSKLLAVVAQEGSSNRAKIDDTGDTVLAYIIDAGSANDA
 SLSQAGYNTMTIQRKSNKANKITQYGTQKTAIVVOROSQAIKRVTR"
 1253..1711
 /gene="csga"
 1253..1711
 /gene="csga"
 /gene="csga"
 /codon_start=1
 /transl_table=1
 /product="Csga protein"
 /protein_id="AAK53212.1"
 /db_xref="GI:14039401"
 /translation="MKLLKRYAAIAIVPSSGALAGVVPYQGGGCGNHGGGAGNNSGPNS
 ELNTVYGGGNSALALQADARNSDLTITQHGCGAGVAGGGSDSSIDLTQGGFGNSA
 TLDQWNGKDSHMTVKQFGGCGAALAVDQTAISNTVNTQVFGNNATAHGY"
 1253..1711

Query Match	49.48;	Score 225.4;	DB 1;	Length 1711;
Best Local Similarity	69.78;	Pred. No. 3.8e-57;		
Matches 320;	Conservative 0;	Mismatches 157;		

QY 1 ATGAAACTTTTAAAGTGGCAGCAATGCGACCAATGTAAGTTTGTGGCAGTGTCTGGCT 60
 Db 1253 ATGAAACTTTTAAAGTAGCAGCAATGCGACCAATGTAATTTCTCCGTAGCGCTGGCA 131
 QY 61 GGCCTGTATGATCGAGCTGTACCCGTGTGTAAACCA--TGAAATGGCAATGCATCG 117
 Db 1313 GGATGTGTTCTCAGTAGCGGGGGGGGTACCACGGTGTGGCGGTAAATACAGC 137
 QY 118 GGCCTGACTCAACGTTGACATTATACGTACGTTCCGCTAAGCTTCCGTTCTCTG 177
 Db 1373 GGCCTGAATTCAGAGCTGAATATTTATTCAGTAGCGGTGGTAACTTCGCACTTCTCTG 143
 QY 178 CAAGCGATGCCGTAATCTGAAAGACATTATCCAGCGGTTATGTATGTAACGCGCC 237
 Db 1433 CAAGCTATCTCTGTACTCTGATCTTACTATATACCGCATGTGGTGGTAAAGGTGCA 1492
 QY 238 GATGTAGGCGCAGGTGCGGATATATGTAATCTATTGAATGACTCAAGATGGTTTCAGAAAT 297
 Db 1493 GATGTGTGTGACGGGCTCAGATGACAGCTCAATCGATGTGACCCAGCTGGCTTTGGTAAC 1552
 QY 298 AATGCCAATGACACAGTAGAAAGCTTAAATCTCCATATTACTGTGCGCCAAATACGCG 357
 Db 1553 ACGCGCACTCTTGATCACTGTGAACGGTAAACCTCTCAATGACAGTTAAACAATTCGGT 1612
 QY 358 GGTAATTAACGCGCGCTGCGTTAATCAACCGCATCTGATTCAGCGTAAATGGTGGCTGAC 417
 Db 1613 GCGCGCAACGGGTGACGGGTGACCAACATCGCATCTTAATTCACCGTCAACAGTAACCTCAG 1672
 QY 418 GTGTGTTTGGCAACAACGCCACGGCTAACCAAGTAAAT 456
 Db 1673 GTTGCTTTGGTAAACAACGCGACCGCTCAATCACTACTAA 1711

	RESULT 12			
	AE005315			
	LOCUS			
	DEFINITION			
	ACCESSION			
	VERSION			
	KEYWORDS			
	SOURCE			
	ORGANISM			
	REFERENCE			
	AUTHORS			
	AE005315	10190 bp	DNA	linear BCT 21-MAR-200
	Escherichia coli O157:H7 EDL933	genome, contig 1 of 3,		section 139
	of 156			
	AE005315 AE005174			
	AE005315.1 GI:12514572			
	. Escherichia coli O157:H7 EDL933			
	Escherichia coli O157:H7 EDL933			
	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
	Enterobacteriaceae; Escherichia.			
	1 (bases 1 to 10190)			
	Perna R.T., Plunkett, G. III, Burland V., Mau, B., Glaesner, J.D.,			
	Rose, J.G., Mayhew, G.F., Evans, P.S., Gregg, J., Kitchpatrick, H.A.,			
	Groebek, E.J., Hackett, J., Klinc, S., Boutin, A., Shao, Y., Miller, L.,			
	Apodaca, J., Anantharaman, T.S., Lin, J., Yen, G., Schwartz, D.C.,			

```

TITLE      Welch R.A. and Blattner F.R.
JOURNAL    Genome sequence of enterohemorrhagic Escherichia coli O157:H7
MEDLINE    Nature 409 (6819), 529-533 (2001)
PUBMED     21074935
AUTHORS    2 (bases 1 to 10190)
            Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
            Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
            Goebel G., Hackett J., Klink S., Boutin A., Shao X., Miller L.,
            Grobeck E.J., Davis N.W., Lim A., Dimalanta E., Potamousis K.,
            Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
            Welch R.A. and Blattner F.R.
TITLE      Direct Submission
JOURNAL    Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES   Location/Qualifiers
            1..10190
            /organism="Escherichia coli O157:H7 EDL933"
            /mol_type="genomic DNA"
            /strain="EDL933"
            /serotype="O157:H7"
            /db_xref="taxon:155864"
            /note="enterohemorrhagic"
            617..1072
            /gene="csgB"
            /note="synonym: Z1675"
            617..1072
            /gene="csgA"
            /function="structural component; Cell exterior
            constituents: Surface structures"
            /note="Residues 1 to 151 of 151 are 100.00 pct identical
            to residues 1 to 151 of 151 from Escherichia coli K-12
            Strain MG1655: B1041"
            /codon_start=1
            /transl_table=1
            /product="minor curlin subunit precursor, similar to CsgA"
            /db_xref="GI:12514573"
            /protein_id="AAG55787.1"
            /translation="MKKKLPMITIGAPGIAAAGYDLANSEYNAVNELSKSEN
            QALITGGATNNSAQLRQGSKLVAAGSSRRAKIDQGDYNLAITDAGSANDA
            SISQAGNTAMIIQKSGNKANITQGTOKTALIVORSOAIVTOR"
            1113..1571
            /gene="csgA"
            /note="synonym: Z1676"
            1113..1571
            /gene="csgA"
            /function="regulator; Cell exterior constituents: Surface
            structures"
            /note="Residues 1 to 152 of 152 are 96.71 pct identical to
            residue 1 to 151 of 151 from Escherichia coli K-12 Strain
            MG1655: B1042"
            /codon_start=1
            /transl_table=1
            /product="curlin major subunit, coiled surface structures;
            cryptic"
            /protein_id="AAG55788.1"
            /db_xref="GI:12514574"
            /translation="MKLLKVAIAIAIVFSGSALAGVVPQGGGNGGGGNGGSPNS
            ELNIVYGGNSALALQADANSLDITTOHGGGAGVGGSDSDSIDLTOHGGNSA
            TLDQNGKDSHMTVKVQGGNGAIVDQASNSTVNTVQVGNATAHQV"
            1630..1962
            /gene="csgC"
            /note="synonym: Z1677"
            1630..1962
            /gene="csgC"
            /function="putative structure; Not classified"
            /note="Residues 1 to 110 of 110 are 98.18 pct identical to
            residues 1 to 110 of 110 from Escherichia coli K-12 Strain
            MG1655: B1043"
            /codon_start=1
            /transl_table=1
            /product="putative curli production protein"
            /protein_id="AAG55789.1"

            /db_xref="GI:12514575"
            /translation="MNALLILAAASSQITNTTQGGDMTITPEVTLTQSLCQVOLL
            SLRSGSGSQSTQKQKTLSPANQPIALTKLISIPDDRKIVIVTSDGSLHSQQ
            WPSSEKS"
            2083..2394
            /gene="z1678"
            2083..2394
            /gene="z1678"
            /function="orf; Unknown function"
            /note="Residues 1 to 103 of 103 are 98.05 pct identical to
            residues 1 to 103 of 103 from Escherichia coli K-12 Strain
            MG1655: B1044"
            /codon_start=1
            /transl_table=1
            /product="orf; hypothetical protein"
            /protein_id="AAG55790.1"
            /db_xref="GI:12514576"
            /translation="MPPPLNLSLGLFPFPIAAGTAQGVIIHFQIVPACDV
            STGSSPVEMNCPOGSGVPGKTYSSKALMSGVNRQAQIASVVOYLDKQKLAWNIEY
            N"
            2489..3022
            /gene="z1679"
            2489..3022
            /gene="z1679"
            /function="orf; Unknown function"
            /note="Residues 1 to 177 of 177 are 100.00 pct identical
            to residues 1 to 177 of 177 from Escherichia coli K-12
            Strain MG1655: B1045"
            /codon_start=1
            /transl_table=1
            /product="putative polyprotein"
            /protein_id="AAG55791.1"
            /db_xref="GI:12514577"
            /translation="MKTRHIVQGDITKLAVDIVANANPSLMGGGVGAIHRAAP
            ALDCAKLVQOQOQDCEPTGHAVITLADLPKAAVAVTVGVMRGEQNEQDLDAYL
            NSRLIVANYSIVAPPAISGVYGVRAAAAEIAVTVSEFTIRHALPEQYVVCVD
            ENNAHLVERLLTQGGDS"
            2964..4445
            /gene="ymcC"
            /note="synonym: Z1680"
            2964..4445
            /gene="ymcC"
            /function="putative enzyme; Not classified"
            /note="Residues 1 to 493 of 493 are 99.59 pct identical to
            residues 1 to 493 of 493 from Escherichia coli K-12 Strain
            MG1655: B1046"
            /codon_start=1
            /transl_table=1
            /product="putative synthase"
            /protein_id="AAG55792.1"
            /db_xref="GI:12514578"
            /translation="MMKKPTSTKSDLPNKEMNDLPRLASAVLPSCQHPQCGFLPL
            EKSLLDAFARYRLAEMSEHTLDVQYIMQDDMGRLFLSALLAAKGVVRLLDND
            NTPGLDILRLLSHPRIEVRLEFNPSEFRLPLRGYITDSRLRRRANKSFYDGVV
            TIVGRIIDGAVPGAGEEPLFSDLVNAIPVBDVADPARVWYCKSVPLQGVLV
            PEGKADRILELPASWHDANTHRYLRMESSPTNHWVDGTLPIWAKTKLLSDPAK
            GEGAKRSHSLPQRLFDIKGSPSERIDIISSYFVPTAGVAQLRWKTKVIAITLN
            SLAANDAVAVHAGVARRKRLRYGVELYELKEPTBQSSYLAHRIQNGSGASIAHT
            FSDIKGVFPGSFNFPSTLNTMEMFVLESTLAQIDKRFIOGQYDAWAQWLDR
            WGRINWDRAKKEIYVKEPATSPFMRVWVRLASILPVEMLL"
            complement(4453..5610)
            /gene="ymdD"
            /note="synonym: Z1681"
            complement(4453..5610)
            /gene="ymdD"
            /function="orf; Unknown function"
            /note="Residues 1 to 385 of 385 are 99.22 pct identical to
            residues 1 to 385 of 385 from Escherichia coli K-12 Strain
            MG1655: B1047"
            /codon_start=1
            /transl_table=1
            /product="orf; hypothetical protein"
            /protein_id="AAG55793.1"

```


gene
150. .2258
/db_xref="taxon:83334"
/gene="EC81032"
150. .2258
/gene="EC81032"
/note="similar to YchY ECOLI gi|1787181 percent identity 99 in 702 aa (conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative oxidoreductase"
/protein_id="BAB34455.1"
/db_xref="GI:13360492"
/translation="WNSLPASTARGLEELLETLENIGAVECQVQGVVHFKGDTRLVYQSILMSLASRIMPLGECVYSDDLVDLYVOAINTWTEPNPATFVAHFSGLNDTIRNSQYGAKKVDAIVDATFRKLPRPNVDRAPDRIANVWHLKETAIDLSDGDLHSGYGGAKVADIPKETTALAAIVMSGQPGPTPLDPMGSGSTGLIEAAMATORAPGLHGRGSPFGMKHDEAIIWQVKACQTPARGGLAEYSRHPGSDSARVIOARTNARLAGIELTFPEYKVAULTNPPLPKPYTVALSNPPYGRDSEPALIALSHLTIHKNOFGEMNLSLFSASPDLLSCQLRADQYKAKNPPLCCVONVHASTPDSKRAMVAEDYARLKRLKKPEKARQGEIECYRLVDYDALEZYNAVDYADWVVOVEYAPKTI DAHKARQLPDIATISVLGIAIPKVLITKRECKQKNOYOKGEEGPELEZYNAHLMVNLTVLDGLFLDIRIARMAPGMSKKDPLNIPSYGSAITVAAGIGARSTTVDMSTTEMAERPLRLNGLTGRPAHRLIQADCLMALEANEQPLITIDPTSSNSGRMEDADVDORHIALMKDLKRLIRAGGITMISNRRGFRMDGLAKGLKRAEITQKTLSDQFARRQKHNCLITAA"

gene
2270. .4177
/gene="EC81033"
2270. .4177
/gene="EC81033"
/note="similar to UDP ECOLI gi|1787182 percent identity 99 in 635 aa (conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative ATP-binding component of a transport system"
/protein_id="BAB34456.1"
/db_xref="GI:13360493"
/translation="MSLISMGWMLSPSDAPRLDMAELIEDNERVCLVGRNAGKSTLMKLNIRGQIDGRIYEDDLIVARLPDPRVSEYVDFAEGIEBOAYLKYHDISRLVMDPSEKUNELAKQEOQLDHNHMOLENRINEVLAQGLDPVAVHSLSGWLKRALQALVSNPVLVLDPEPTNHLDIETDWLBSGLFTNGTILFISHDSFLIKMATRIVLDLRKLVTVPLGPNYDOYLFVEKEALRVBELQNAEFTKQAEVWIIQGIKARSTRNEGVAIKAMRREGRREVMGTAKKQVEASRSGKIPEMEDVCYVDGKLVKDSFAOVLRGDKIALIGPNCGKTTLLKMLGOLADSGRINHGTLEVAYDQHRAELDPKTYMDMLASGKQEVWNGKPRHYLTGLDPLPHPRANTPPRALSGERNRLLIARLKRSLNLIIDEPNDLDVEFTLELIDSYGTVLAVSHDQFADNTVTEGMIFEGGKIGRVYGHDKARQQEQYVALKQVPAKVAETVRSKSLTYCLORELEQLQLLEDLEAKLEALQTVADASFPSQHEQTKVLADMAAEOELBOAERMEYLEALNNGG"

gene
4307. .5560
/gene="EC81034"
4307. .5560
/gene="EC81034"
/note="similar to POJA ECOLI gi|1787183 percent identity 99 in 417 aa (conserved in E.coli K-12)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="paraquat-inducible protein A"
/protein_id="BAB34457.1"
/db_xref="GI:13360494"
/translation="MCEHHNAKHILISQCDMLVVALPRLHGOAKACPRGTTLYTAWDAPRPRAYALALFMILLSNLPFPNMNVAVGTSTITLIEPGVLFSDVASLGFVFLPQVPAACLTITLLILNRAELPRLKEQLARVLFOLKTGMEALIFLAGVASFVKLMAVGSIGVSSSFLPMCLFCVLQIARFOCDRMDLMDIAPMBELQPLKPCVTRIGOLNAGSCCTAIIIPADEBPCRCGTGKGVARRNSLOTLALIVTSIMLYPANILPIMVTDLGSGMPTIAGVILVMSSEGSYPAAVITPLASIMVPLTKMIATMCMKAGHGKDSRPMILYIEVIVFGWRMSMDIVFVAVIALVVRGGLSMSTYIPMAGMLPMLVYIMTMPSMTDPRPLSDMDROPSEBHS"

gene
5565. .7205
/gene="EC81035"

5565. .7205
 /gene="ECs1035"
 /note="similar to POIB_ECOLI gi|1787184 percent identity
 99 in 546 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="paraquat-inducible protein B"
 /protein_id="BAB34458.1"
 /db_xref="GI:13360495"
 /translation="MESNNGEAKIOKVKVMSPVWIPPIVATLIGAWVLPHYSHQPE
 VTLITRNABGIEGCKTTIKTSRVDVSVSATIADLTLYEIKARLNSGNEKLKHRT
 VFWVRKPIGREGISGLTLLSGVITELDPGAGSKMDKDLSDPELAPDPAKGIIV
 VLDSKAGQLSPGDVPLFRGYVGVSVETSPFDQKNIISYQLFINAEYDLVTSNVE
 WMDSGIAVDLTISAGKRVEMSGJLTLLSGVSPFVPEGLDGGVPAKTAFLVYDDGS
 IODSLYTHIDIKMFLPFKDSVRGLQCPAPFPGIRLGTSGSKVFPFPMNMQTFNDQPS
 IPIVLRIPERIKMFLQMGADVVEHAGELIKRGLRSLITGQNLVTCALVYDLDFPPT
 PAITIRERNYQIILPTVSGGLAQIQRLMEALDKINKPLPNDMTCQSTLSRESORT
 MKNLEDTLPVLTKNIILASOSMQLPTDMOSTLEELNWSGQFPGSAAYNRMVADMRL
 DQNELDLPVLTKNIILASOSMQLPTDMOSTLEELNWSGQFPGSAAYNRMVADMRL
 7217. .7765
 /gene="ECs1036"
 7217. 7765
 /gene="ECs1036"
 /note="similar to YMB_ECOLI gi|1787185 percent identity
 98 in 182 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAB34459.1"
 /db_xref="GI:13360496"
 /translation="MTIALMTLACSGSGEINKRNVQLPVQSGTOSTASGNLLWVE
 QVAVDYLAGNVGVYQSDVKYVIANNMAMAPLDQLNTLIVANLSTOLPGNVVLSO
 PLGSGQDTLVNTVTEFRNGRYDCKVIVSGEMLNHQQLIKRPRLSGVQIQDGYDENV
 KVLAVGWSOEALSIQSEIRLPL
 8021. .8188
 /gene="ECs1037"
 8021. 8188
 /gene="ECs1037"
 /note="identical to RMF_ECOLI gi|1787186 (Conserved in
 E.coli K-12)"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11
 /product="ribosome modulation factor"
 /protein_id="BAB34460.1"
 /db_xref="GI:13360497"
 /translation="MKRQKRDRLERAHQRYQAGIAGRSKEMCPYOTLNORSOMLGSM
 REAMADRVVMA"
 complement(8258. .8776)
 /gene="ECs1038"
 complement(8258. .8776)
 /gene="ECs1038"
 /note="similar to FABA_ECOLI gi|1787187 percent identity
 100 in 172 aa (Conserved in E.coli K-12)"
 /codon_start=1
 /evidence=not experimental
 /transl_table=11

```

/codon_start=1
/transl_table=11
/product="assembly /transport component in curli
production"
/protein_id="CAA62277.1"
/db_xref="GI:11479559"
/db_xref="CDB:P52103"
/db_xref="SMISS-PROT:P52103"
/transeq_id="MORLELVAWMLSGCLTAPPEKARPTLMPPASQYKLTPLP
PYGKIPSYVNIODETGQCPKPASNESTAPAQAATMLTALKSRMRPIITLERGL
NLAERKILPAAGENGTAANNRIPLQSITZANIMEGSSIGESNVKSGGARYPRF
ICGDHIOLODOLAVNLRLVNNVSTGEIISVTSFTLLSYEVONAGRPFIIDIRLBEGS
VGTTSNBPMLCMIAETGVIFLINDIGRLMDIQNKAKERNQDLIVKIKRMSVPPE
S"
/complement(993..1409)
/gene="csgF"
/complement(993..1409)
/gene="csgF"
/codon_start=1
/transl_table=11
/product="assembly /transpoct component in curli
production"
/protein_id="CAA62277.1"
/db_xref="GI:11479559"
/db_xref="CDB:P52103"
/db_xref="SMISS-PROT:P52103"
/transeq_id="MORLELVAWMLSGCLTAPPEKARPTLMPPASQYKLTPLP
PYGKIPSYVNIODETGQCPKPASNESTAPAQAATMLTALKSRMRPIITLERGL
NLAERKILPAAGENGTAANNRIPLQSITZANIMEGSSIGESNVKSGGARYPRF
ICGDHIOLODOLAVNLRLVNNVSTGEIISVTSFTLLSYEVONAGRPFIIDIRLBEGS
VGTTSNBPMLCMIAETGVIFLINDIGRLMDIQNKAKERNQDLIVKIKRMSVPPE
S"
/complement(993..1409)
/gene="csgF"

```

```

gene
    complement (993. .1409)
    /gene="csgF"
    complement (993. .1409)
    /gene="csgF"
    /codon_start=1
    /transl_table=11
    /product="assembly"
    /transport_component in curl1
    /protein_id="CAA62278.1"
    /db_xref="GI:1147560"
    /translation="MRVKAIVVILMIISPLISAGMTFORENPFGGNPNNGAFILNS
AQAQNSKXDSDYNDVDFEIGETPSALNDFQDAIQSLIGGLSLININTGPKGMTVDYIV
DIARNDQLOLVNDRKGTGTSITVSGELQNNSTDP"
    complement (1434. .1823)
    /gene="csgE"
    complement (1434. .1823)
    /gene="csgF"

```

```

/ trans1_table=11
/ product="assembly / transport component in curli
production"
/ protein_id="CAA62279.1"
/ db_xref="GI:11476161"
/ db_xref="SWISS-PROT:P52105"
/ translation="MKRIILKIVTAIEFLFAAGILHAVEVEVPGILTDHTVSISCHDHY
IQRFEELNRQIQNLSTSDLAHDEF"
complement(1828..2478)
/ gene="csgd"
/ complement(1828..2478)
/ gene="csgd"
/ note="necessary for transcription of the csbA operon"
/ codon_start=1

```

```

/misc_feature
    /note="transcriptional start site for csdDEG operon"
    3140
/misc_feature
    /note="transcriptional start site for csdDEG operon"
    3233..3688
/misc_feature
    /gene="csdG"
    3233..3688
/misc_feature
    /gene="csdG"
    /codon_start=1
    /transl_table=1
    /product="nucleation component of curlin monomers"
    /protein_id="CA62281.1"
    /db_xref="GI:1147563"

```

gene
CDS
/translation="MKKLLFMWLTILGAPGIAAAGYDLANSEYNFAPANELSKSSFN
QAATIGAGCTNNASQLROGSKSLAVNAGESSNAPKIDQGDYHATLATIDAGSANDA
SISGAYCNTAMTIIOKSGSKNANTIOYGIKATVVOHOSQMAIKVOTR"
3729..4184
/gene="csgA"
3729..4184
/gene="csgA"
/note="major subunit"

gene
CDS
/translation="MKLKVAAIAIAIVSGSALAGVPOYGGGNGHGGNNSGPNS
LDNIYOGGNSALAQTDARNSDLTIHOGGNGADVGQSDSDSILDTQSGFGNSAT
LQWNGKSEMTVQFGGGNGAAVQDTASNSVNVTVQFGNNATAHQY"
3729..3788
/gene="csgA"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/transl_table=11
/product="putative curli production protein"
/protein_id="CAA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLLAALSSQITFTNTQGGDYVITIIPEVTLTQSLCVOIL
SLRSGSGSOTKQKLTSLPANQPIALTKLSINISPDVKYIVTVTSDDGSLHLSQO
WPSSEKX"

sig_peptide
gene
CDS
/translation="MKLKVAAIAIAIVSGSALAGVPOYGGGNGHGGNNSGPNS
LDNIYOGGNSALAQTDARNSDLTIHOGGNGADVGQSDSDSILDTQSGFGNSAT
LQWNGKSEMTVQFGGGNGAAVQDTASNSVNVTVQFGNNATAHQY"
3729..3788
/gene="csgA"
4243..4575
/gene="orfC"
4243..4575
/gene="orfC"
/codon_start=1
/transl_table=11
/product="putative curli production protein"
/protein_id="CAA62283.1"
/db_xref="GI:1147565"
/db_xref="SWISS-PROT:P52107"
/translation="MNTLLLAALSSQITFTNTQGGDYVITIIPEVTLTQSLCVOIL
SLRSGSGSOTKQKLTSLPANQPIALTKLSINISPDVKYIVTVTSDDGSLHLSQO
WPSSEKX"

Query Match 49.1%; Score 224; DB 1; Length 4680;
Best Local Similarity 68.2%; Pred. No. 1.1e-56;
Matches 311; Conservative 0; Mismatches 145; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGCGAGTCTTGCT 60
DB 3729 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGCGAGTCTTGCT 3788
QY 61 GGCGCTATGATCACTGCTGTTACCGGTGTGTGTAACCAATGACATGATCGGC 120
DB 3789 GGCGCTATGATCACTGCTGTTACCGGTGTGTGTAACCAATGATAGCGGC 3848
QY 121 CCGGACTCAAGTTCGATTCAGTACGCTTCGCTAAGCGTGGCTGCTCTGCA 180
DB 3849 CCAATTCGATTCAGTACGCTTCGCTAAGCGTGGCTGCTGCACTGCTCTGCA 3908
QY 181 AGCGATGCGCTTAAATCTGAAACGACATTCACGACGCGTTATGTAACGCGCGCAT 240
DB 3909 ACTATGCGCGCTTAACTGACTGATTAACCGACGATGCGCGGTTATGTAACGAT 3968
QY 241 GTAGCGGCGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
DB 3969 GTAGCGGCGGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 4028
QY 301 GCCACATCGACGATGGAAGCTTAAATCCGATTTATCTGCGCAATTAAGCGCGGT 360
DB 4029 GCTACTCTGATCACTGGAACGCAAAATCTGAAATGACGCTTAAACATCTGCGGT 4088
QY 361 AATAACGCGCGGTGTTAATCAACCGCATCTGATTCGACGTAATGTCGTCAGTT 420
DB 4089 GCGAAGCGGTGTCGATGACCAAGCTGATTAATCTCTCGTCAAGTACAGGT 4148
QY 421 GGTTCGCAACAGCGCGCTAATCAAGATTA 456
DB 4149 GGTTCGCAACAGCGCGCTAATCAAGATTA 4184

RESULT 15

AE000205
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
SOURCE
1.10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"

```

repeat_region complement(<1. .48)
promoter /note="IS3" .
promoter /note="factor Sigma70; predicted +1 start at 1094689"
gene /note="factor Sigma70; predicted +1 start at 1094716"
CDS /gene="b1028"
/gene="b1028"
/function="orf; Unknown"
/note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
to 58 residues of an approx. 188 aa protein RAS2_CANAL
SW: P43098"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/db_xref="GI:178726"
/translation="MEDGVQARDYDLGLDIASPEHOVLNNYAKSEVAPSTIKENLS
LHHTVKTESLDTFLNMPVPAFLISNKPYLGNMLNDFAGVDQQRVHCKKAFSD
AKVFK"
275. .299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
389. .1375
/gene="ycdu"
/note="synonym: b1029"
389. .1375
/gene="ycdu"
/function="orf; Unknown"
/note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
to 191 residues of an approx. 288 aa protein Y320_MYCGE
SW: P47562"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIPYLPIRPODKRNIIYYAIGLIGFYMNAGTPPSEDI
LIMLSMWGLYOLPDIKAYAKVCVKNIIDFSMPKKONHKELEFLQPLVAGTFS
SLRVYLSVCISTYVNLITISPLNDSFVTEGOMLPKSIITALLIGCVATNLF
PLRPSKRIAFGRFLQIDLFESSENLSTFPKPLMLRFLITLVETWMTLVSVL
ALVLEKRIFFEYFLCYVPCLIYFYFCRPLMNDPMACDMYFRKGFHK"
1253. .1280
/gene="ycdu"
/note="central position to predicted promoter: -104"
1253. .1280
/bound_moiety="Deor predicted site"
/gene="ycdu"
/note="No predicted promoter"
1337. .1365
/bound_moiety="Deor predicted site"
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494. .1745
/gene="b1030"
/gene="b1030"
/function="orf; Unknown"
/note="083; This 83 aa ORF is 25 pct identical (7 gaps)
to 81 residues of an approx. 616 aa protein ALBU_RABIT
SW: P49065"
/codon_start=1
/transl_table=1
/product="orf; hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGRVYOTKKKPPUSYELFRYGGEGIDLSRSPCGQPAHC
VHGLSNLSPVDPSPRSGLLIPVCAIHEKKAITVRL"
1513. .1926
/gene="ycdu"

```

```

CDS /note="synonym: b1031"
1513. .1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="0137; This 137 aa ORF is 27 pct identical (1 gap)
to 47 residues of an approx. 80 aa protein RS3_YEAST
SW: P05759"
/codon_start=1
/transl_table=1
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSPYRTSSSLMAVRGLTRPAPAGSLTAFTYCP
GKROLSPGRGESSRVCQVYTKKPYLSYELFPPKGGGIDLSRSPCGAHCVHGL
SNLSPYBRSGLIPVCAIYEKKARFPVAL"
complement(2111. .2198)
/gene="serX"
/note="synonym: b1032"
complement(2111. .2198)
/gene="serX"
/product="tRNA-Ser"
/note="anticodon: GCA; CG Site No. 17878"
/function="tRNA; tRNA"
2269. .2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338. .2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355. .2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393. .3370
/gene="ycdu"
/note="synonym: b1033"
2393. .3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="0325; This 325 aa ORF is 32 pct identical (2 gaps)
to 177 residues of an approx. 312 aa protein YPBA_CORGL
SW: P45637"
/transl_table=1
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="KRIKFAQFSGVSDIIFYHPTPTOMWIEALRKAIPOARVAM
KSGDNDADYALVWHPPEVMEAGRDIAVFALGAGVDSILSKLAHPMLNPSPLR
LEDTGMEQOEYAVSOVLHMRFRPDYRIQONSHQIPDYREPTTIGIDAGVL
GSKVAQSLQTMRFPLRCWSRTERKSPGWQSGRGRBSAPLSQCRVLINLPETV
GINOOLEKLEPDGAYLNLARGVYVDDLLALDGRVAGALDVFNRPPLPESP
LMQHPVITTPHVAITRPAVAVEYISTTIAQLKEKRGVCGVDARAGY"
3278. .3308
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1097992"
3383. .3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396. .3423
promoter
promoter
promoter
Query Match 49.1%; Score 224; DB 1; Length 10346;
Best Local Similarity 68.2%; Pred. No. 1.2e-56;
Matches 311; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGAGCGACATTCGACGCAATCGTAGTTCGCGAGTCGTGGCT 60
DB 8993 ATGAACCTTTAAAGAGCGACATTCGACGCAATCGTAGTTCGCGAGTCGTGGCT 60
QY 61 GAGCGTATGATCGAGCGGTACCCGCTGTTTATCCATGAATGCGACATCGCGC 120
DB 9053 GATGTTGTTCTCGATACGGCGGGGAGAACCAAGGTGGCGGTAAATATGCGCGC 9112
QY 121 CCGACCTCAAGTTGACATTTATCAGTACGCTCCGCTAACGCGCGCTTCTTGCA 180
DB 9113 CCAATTTGAGCTGAACATTTACAGTACGCTGGCGGTAACTTGCACTTCTTGCA 9172
QY 181 AGCGATCCCGTAATCTGAAGACATTAACCCAGAGCGGTATGTATGAGCGCGCAT 240

```

	9173	9174	9175	9176	9177	9178	9179	9180	9181	9182	9183	9184	9185	9186	9187	9188	9189	9190	9191	9192	9193	9194	9195	9196	9197	9198	9199	9200	9201	9202	9203	9204	9205	9206	9207	9208	9209	9210	9211	9212	9213	9214	9215	9216	9217	9218	9219	9220	9221	9222	9223	9224	9225	9226	9227	9228	9229	9230	9231	9232	9233	9234	9235	9236	9237	9238	9239	9240	9241	9242	9243	9244	9245	9246	9247	9248	9249	9250	9251	9252	9253	9254	9255	9256	9257	9258	9259	9260	9261	9262	9263	9264	9265	9266	9267	9268	9269	9270	9271	9272	9273	9274	9275	9276	9277	9278	9279	9280	9281	9282	9283	9284	9285	9286	9287	9288	9289	9290	9291	9292	9293	9294	9295	9296	9297	9298	9299	9300	9301	9302	9303	9304	9305	9306	9307	9308	9309	9310	9311	9312	9313	9314	9315	9316	9317	9318	9319	9320	9321	9322	9323	9324	9325	9326	9327	9328	9329	9330	9331	9332	9333	9334	9335	9336	9337	9338	9339	9340	9341	9342	9343	9344	9345	9346	9347	9348	9349	9350	9351	9352	9353	9354	9355	9356	9357	9358	9359	9360	9361	9362	9363	9364	9365	9366	9367	9368	9369	9370	9371	9372	9373	9374	9375	9376	9377	9378	9379	9380	9381	9382	9383	9384	9385	9386	9387	9388	9389	9390	9391	9392	9393	9394	9395	9396	9397	9398	9399	9400	9401	9402	9403	9404	9405	9406	9407	9408	9409	9410	9411	9412	9413	9414	9415	9416	9417	9418	9419	9420	9421	9422	9423	9424	9425	9426	9427	9428	9429	9430	9431	9432	9433	9434	9435	9436	9437	9438	9439	9440	9441	9442	9443	9444	9445	9446	9447	9448	9449	9450	9451	9452	9453	9454	9455	9456	9457	9458	9459	9460	9461	9462	9463	9464	9465	9466	9467	9468	9469	9470	9471	9472	9473	9474	9475	9476	9477	9478	9479	9480	9481	9482	9483	9484	9485	9486	9487	9488	9489	9490	9491	9492	9493	9494	9495	9496	9497	9498	9499	9500	9501	9502	9503	9504	9505	9506	9507	9508	9509	9510	9511	9512	9513	9514	9515	9516	9517	9518	9519	9520	9521	9522	9523	9524	9525	9526	9527	9528	9529	9530	9531	9532	9533	9534	9535	9536	9537	9538	9539	9540	9541	9542	9543	9544	9545	9546	9547	9548	9549	9550	9551	9552	9553	9554	9555	9556	9557	9558	9559	9560	9561	9562	9563	9564	9565	9566	9567	9568	9569	9570	9571	9572	9573	9574	9575	9576	9577	9578	9579	9580	9581	9582	9583	9584	9585	9586	9587	9588	9589	9590	9591	9592	9593	9594	9595	9596	9597	9598	9599	9600	9601	9602	9603	9604	9605	9606	9607	9608	9609	9610	9611	9612	9613	9614	9615	9616	9617	9618	9619	9620	9621	9622	9623	9624	9625	9
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

Search completed: March 15, 2004, 22:50:09
Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagcgc.....ccacgcctaccagtattaa 456.

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3	AAC64624 Agfa::PT3
2	403.2	88.4	456	2	AAQ87467 Agfa sequ
3	403.2	88.4	456	2	AAAT74142 Salmonell
4	403.2	88.4	456	3	AAAC64617 Salmonell
5	360	78.9	456	3	AAAC64626 Agfa::PT3
6	353.6	77.5	456	3	AAAC64625 Agfa::PT3
7	352	77.2	456	3	AAAC64628 Agfa::PT3
8	352	77.2	456	3	AAAC64622 Agfa::PT3
9	352	77.2	456	3	AAAC64629 Agfa::PT3
10	352	77.2	456	3	AAAC64623 Agfa::PT3
11	348.8	76.5	456	3	AAAC64630 Agfa::PT3
12	348.8	76.5	456	3	AAAC64627 Agfa::PT3
13	345.6	75.8	456	3	AAAC64631 Agfa::PT3
14	284.2	62.3	361	2	AAQ73066 Agfa sequ
15	284.2	62.3	361	2	AAAT74141 Salmonell
16	222.4	48.8	456	3	AAAC64619 Escherich
17	222.4	48.8	456	9	AAAC64619 Escherich
18	210	46.1	646	2	AAAC64617 Salmonell
19	149.6	32.8	359	2	AAAC64617 Salmonell
20	51.2	11.2	100	7	AAAC64617 Salmonell
21	49	10.7	78	9	AAAC64617 Salmonell
22	48.2	10.6	78	9	AAAC64617 Salmonell
23	48	10.5	48	3	AAAC64621 Salmonell

24	48	10.5	78	3	AAAC64610 Agfa (SEF
25	43.2	9.5	48	3	AAAC64615 E. coli K
26	42.4	9.3	100	7	AAAC64615 E. coli K
27	41.6	9.1	72	9	AAAC64615 E. coli K
28	41.2	9.0	78	3	AAAC64605 Sefta (SEF
29	40.8	8.9	100	7	AAAC64605 Sefta (SEF
30	40.6	8.9	78	3	AAAC64606 Sefta (SEF
31	35.6	7.8	1965	4	AAAC64627 Agfa::PT3
32	35.6	7.8	1965	7	AAAC64627 Agfa::PT3
33	35.4	7.8	3411	5	AAAC64620 Escherich
34	35.4	7.6	456	3	AAAC64620 Escherich
35	34.6	7.6	1506	6	AAAC64618 Salmonell
36	34.6	7.6	456	3	AAAC64618 Salmonell
37	34	7.5	2000	7	AAAC64618 Salmonell
38	34	7.5	2583	7	AAAC64618 Salmonell
39	34	7.5	2583	7	AAAC64618 Salmonell
40	34	7.5	2886	7	AAAC64618 Salmonell
41	34	7.5	3300	3	AAAC64618 Salmonell
42	33.6	7.4	549	3	AAAC64618 Salmonell
43	33.4	7.3	3990	2	AAAC64618 Salmonell
44	33.4	7.3	5100	2	AAAC64618 Salmonell
45	33.4	7.3	7766	2	AAAC64618 Salmonell

ALIGNMENTS

RESULT 1	AAAC64624 standard; DNA; 456 BP.
ID	AAAC64624
XX	AAAC64624;
AC	AAAC64624;
XX	26-FEB-2001 (first entry)
DT	26-FEB-2001 (first entry)
DE	Agfa::PT3#3 DNA sequence SEQ ID NO:15.
XX	Agfa::PT3#3 DNA sequence SEQ ID NO:15.
KW	Salmonella; agfa; chromosomal gene replacement; fimbriin; epitope;
KV	vaccine; immune response; immunogen; ds.
XX	Salmonella enteritidis.
OS	Salmonella enteritidis.
OS	Escherichia coli.
OS	Synthetic.
XX	WO200060102-A2.
PM	12-OCT-2000.
XX	12-OCT-2000.
PD	05-APR-2000; 2000WO-CA000356.
XX	05-APR-2000; 2000WO-CA000356.
PF	05-APR-1999; 99US-0127888P.
XX	05-APR-1999; 99US-0127888P.
PR	(UUVI-) UNIV VICTORIA.
XX	(UUVI-) UNIV VICTORIA.
PA	White AP, Doran JL, Collison SK, Kay WW;
XX	White AP, Doran JL, Collison SK, Kay WW;
XX	WPI, 2000-672631/65.
DR	P-PSDB; AAB36348.
XX	P-PSDB; AAB36348.
XX	Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa
PT	protein useful for eliciting immune response in animal.
PT	protein useful for eliciting immune response in animal.
XX	Disclosure; Page 136; 139pp; English.
PS	Disclosure; Page 136; 139pp; English.
XX	The present invention describes a recombinant agfa gene (1) where a
XX	segment of the gene has been replaced by a segment of a foreign DNA
CC	sequence which encodes a foreign epitope or antigen. Also described are:
CC	(1) use of thin aggregative fimbriae (SEF1/TFP) nucleation depended
CC	assembly system of strains of Salmonella, Escherichia coli and
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant
CC	Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)
CC	directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrin subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 8.5e-141;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTGGCT 60
QY 61 GGGCTCATGATAGCTGGTAAACCGCTGTTGTTACCATGAAATGGCAATGATCCGGC 120
DB 61 GGGCTCATGATAGCTGGTAAACCGCTGTTGTTACCATGAAATGGCAATGATCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGTAACGCTGGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGTAACGCTGGCTCTGCA 180
QY 181 AGCGATGCCCGTAATCTGAAACGACATTCACCGAGCGGTTATGTTACGGCGCAT 240
DB 181 AGCGATGCCCGTAATCTGAAACGACATTCACCGAGCGGTTATGTTACGGCGCAT 240
QY 241 GTAGGCCAGGTCGCGATTAATAGTACTTGTGAATGAGTCAATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGTCGCGATTAATAGTACTTGTGAATGAGTCAATGTTTCAAAATAT 300
QY 301 GCCACATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
DB 301 GCCACATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGTTGCTCAGGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGTTGCTCAGGTT 420
QY 421 GGTTTGGCAACACGCCACGCTTAACCATGATTTAA 456
DB 421 GGTTTGGCAACACGCCACGCTTAACCATGATTTAA 456

```

RESULT 2
AAQ87467
ID AAQ87467 standard; DNA, 456 BP.

AC AAQ87467;
XX
DT 25-MAR-2003 (revised)
DT 26-JUN-1995 (first entry)
XX
DE Agfa sequence.
XX
KM *Salmonella*; Agfa; vaccine; genetic immunization; ds.
XX
OS *Salmonella*.
XX

Key Location/Qualifiers
CDS 1..454
FT /*tag=a
FT /note="Agfa"
PN MO9425598-A2.
XX
XX 10-NOV-1994.
XX
XX 26-APR-1994; 94MO-IB000207.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX (KING/) KING J.
XX
XX Kay WW, Collinson SK, Clouthier SC, Doran JL,
XX WPI; 1994-358275/44.
XX P-PSDB; AAR74625.
XX
XX Eliciting an immune response to *Salmonella* - using attenuated *Salmonella*
XX strains, vector constructs, or comps. contg. fibrillar type proteins.
XX
XX Disclosure; Fig 7B; 95pp; English.
XX
XX The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated
XX proteins are used in genetic immunization and vaccine compositions,
XX respectively, to elicit an immune response to *Salmonella* in animals (e.g.
XX food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN
XX field.)
XX

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 2; Length 456;
Best Local Similarity 92.8%; Pred. No. 2.9e-123;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATGCTGTTCTGGCAGTCTGGCT 60
QY 61 GGGCTCATGATAGCTGGTAAACCGCTGTTGTTACCATGAAATGGCAATGATCCGGC 120
DB 61 GGGCTCATGATAGCTGGTAAACCGCTGTTGTTACCATGAAATGGCAATGATCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGTAACGCTGGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCCGTAACGCTGGCTCTGCA 180
QY 181 AGCGATGCCCGTAATCTGAAACGACATTCACCGAGCGGTTATGTTACGGCGCAT 240
DB 181 AGCGATGCCCGTAATCTGAAACGACATTCACCGAGCGGTTATGTTACGGCGCAT 240
QY 241 GTAGGCCAGGTCGCGATTAATAGTACTTGTGAATGAGTCAATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGTCGCGATTAATAGTACTTGTGAATGAGTCAATGTTTCAAAATAT 300
QY 301 GCCACATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
DB 301 GCCACATGACCAAGTGAACGCTTAAATCTCCATATTAATCTGCGCCAAATACGGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGTTGCTCAGGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATCCAGCGTAATGTTGCTCAGGTT 420
QY 421 GGTTTGGCAACACGCCACGCTTAACCATGATTTAA 456
DB 421 GGTTTGGCAACACGCCACGCTTAACCATGATTTAA 456

```

RESULT 3
AAT74142

ID AAT74142 standard; DNA; 456 BP.
 AC AAT74142;
 XX
 DT 25-MAR-2003 (revised)
 DT 29-SEP-1997 (first entry)
 XX
 DE Salmonella enteritidis 27655-3b agfa gene.
 XX
 KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
 XX
 OS Salmonella enteritidis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..456
 FT /*tag= a
 FT /label= agfa_gene_fragment
 FT /transl_except= (pos:367..369,aa:Pro)
 XX
 PN US563617-A.
 XX
 PD 03-JUN-1997.
 XX
 PP 26-APR-1994; 94US-00233788.
 XX
 PR 26-APR-1993; 93US-00054452.
 XX
 PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
 XX
 PI Collinson SK, Kay WW, Doran JL;
 XX
 XX MPI: 1997-309886/28.
 DR P-PSDB; AAM23570.
 XX
 PT Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
 PT enteropathogenic bacteria of the Enterobacteria family.
 XX
 PS Claim 1; Col 19-112; 85pp; English.
 XX
 CC The present sequence represents an isolated agfa gene derived from
 CC Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
 CC diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
 CC family Enterobacteria. It can also be used to provide proteins and
 CC antibodies which can be used for assays. The nucleic acid sequence can be
 CC used to provide probes or primers which can specifically hybridise to
 CC nucleic acid molecules from greater than 99% of Salmonella strains that
 CC are pathogenic to warm-blooded animals relative to nucleic acid molecules
 CC from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
 CC correct PF field.)
 CC
 SQ Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
 Query Match 88.4%; Score 403.2; DB 2; Length 456;
 Best Local Similarity 92.8%; Pred. No. 2.9e-123;
 Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
 QY 1 ATGAACTTTAAAGTGGACGATCCGACCATCGAGTTCTGGAGTCTTGACT 60
 DB 1 ATGAACTTTAAAGTGGACGATCCGACCATCGAGTTCTGGAGTCTTGACT 60
 QY 61 GCGCTATGATGACGCTGTACCGGTGTGTTACCCATGAATGGCATGATCGCGGC 120
 DB 61 GCGCTATGATGACGCTGTACCGGTGTGTTACCCATGAATGGCATGATCGCGGC 120
 QY 121 CCGAGCTCAAGCTTGAAGCATTATCAGTACGCTTCGGTAAACGCTGCTCTGCA 180
 DB 121 CCGAGCTCAAGCTTGAAGCATTATCAGTACGCTTCGGTAAACGCTGCTCTGCA 180
 QY 181 AGCGATCGCCGCTAATCTGAAACGACCTTACCCAGAGCGGTTATGTAAGCGCGCAT 240
 DB 181 AGCGATCGCCGCTAATCTGAAACGACCTTACCCAGAGCGGTTATGTAAGCGCGCAT 240
 QY 241 GTAGCGCAGGTGGGATATAGTACTATGAACTGACTGAAATGTTTCAAAATAT 300

DB 241 GTAGCGCAGGTGGGATATAGTACTATGAACTGACTGAAATGTTTCAAAATAT 300
 QY 301 GCCACATCGACGATGGAAAGCTAAATCCGATATTTACTGCGGCAATACGCGGT 360
 DB 301 GCCACATCGACGATGGAAAGCTAAATCCGATATTTACTGCGGCAATACGCGGT 360
 QY 361 AATACCGCGGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGCTCAGTT 420
 DB 361 AATACCGCGGCTGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGCTCAGTT 420
 QY 421 GGTTCGCAACCAACGCGCTAACGATTTAA 456
 DB 421 GGTTCGCAACCAACGCGCTAACGATTTAA 456
 RESULT 4
 ID AAC64617 standard; DNA; 456 BP.
 AC AAC64617;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
 XX
 KM Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX
 OS Salmonella enteritidis.
 XX
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PP 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collinson SK, Kay WW;
 XX
 XX MPI: 2000-672631/65.
 DR P-PSDB; AAB36341.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 134; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system or strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising; recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC the directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 88.4%; Score 403.2; DB 3; Length 456;
Best Local Similarity 92.8%; Pred. No. 2,9e-123;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTTGCGAGCATGTTGTTGCGAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTTGCGAGCATGTTGTTGCGAGTCTGCTGCT 60
QY 61 GCGCTCTATGATCAGCTGCTGTTACCGCTGTTGTTACCATGAAATGCGACATCCGCG 120
DB 61 GCGCTCTATGATCAGCTGCTGTTACCGCTGTTGTTACCATGAAATGCGACATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
QY 241 GTAGGCGGAGGTCGAGTAAATAGTACTTATGATGCTGCTGCTGCTGCTGCTGCT 300
DB 241 GTAGGCGGAGGTCGAGTAAATAGTACTTATGATGCTGCTGCTGCTGCTGCTGCT 300
QY 301 GCCACCATCGACAGTGGAGCGTAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 301 GCCACCATCGACAGTGGAGCGTAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 361 AATAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 AATAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 GGTGTTGGCAACACGCGGCTAACCGATTAAT 456
DB 421 GGTGTTGGCAACACGCGGCTAACCGATTAAT 456

```

RESULT 5
AAC64626
ID AAC64626 standard; DNA; 456 BP.
XX
AC AAC64626;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#5 DNA sequence SEQ ID NO:19.
XX
KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX vaccine; immune response; immunogen; ds.
XX
OS Salmonella enteritidis.
XX
OS Escherichia coli.
XX
OS Synthetic.
XX
EN WO200006102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-012788BP.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay MW;

XX WP1: 2000-672631/65.
DR P-PSDB; AAB36350.
XX

Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended
CC assembly system of strains of *Salmonella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong live
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;
Best Local Similarity 86.8%; Pred. No. 6.4e-109;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGAGCATTTGCGAGCATGTTGTTGCGAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTTGCGAGCATGTTGTTGCGAGTCTGCTGCT 60
QY 61 GCGCTCTATGATCAGCTGCTGTTACCGCTGTTGTTACCATGAAATGCGACATCCGCG 120
DB 61 GCGCTCTATGATCAGCTGCTGTTACCGCTGTTGTTACCATGAAATGCGACATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTGCTGCTGCTGCTGCTGCTGCT 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAACGCGCGAT 240
QY 241 GTAGGCGGAGGTCGAGTAAATAGTACTTATGATGCTGCTGCTGCTGCTGCTGCT 300
DB 241 GTAGGCGGAGGTCGAGTAAATAGTACTTATGATGCTGCTGCTGCTGCTGCTGCT 300
QY 301 GCCACCATCGACAGTGGAGCGTAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 360
DB 301 GCCACCATCGACAGTGGAGCGTAAATCTGCTGCTGCTGCTGCTGCTGCTGCT 360
QY 361 AATAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
DB 361 AATAAGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 420
QY 421 GGTGTTGGCAACACGCGGCTAACCGATTAAT 456
DB 421 GGTGTTGGCAACACGCGGCTAACCGATTAAT 456

```

Db 421 GGTITGGCAACAGCCAGGCTAACGATATTA 456

RESULT 6
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
XX AAC64625;
AC
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT#4 DNA sequence SEQ ID NO:17.
DE
XX Salmomella: agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO20060102-A2.
PN
XX 12-OCT-2000.
PD
XX 05-APR-2000; 2000MO-CA000356.
PF
XX 05-APR-1999; 99US-0127888P.
PR
XX (UYVI-) UNIV VICTORIA.
PA
PI White AP, Doran JL, Collison SK, Kay WW;
P1 WPI: 2000-672631/65.
XX P-PSDB; AAB36349.
DR
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/7A) nucleation depended assembly system of strains of Salmomella, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmomella, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbriin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 77.5%; Score 353.6; DB 3; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.5e-107;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGACAGCATTCGTAGTTTCTGGAGTCTTGCT 60
Db 1 ATGAACCTTTAAAGTGGACAGCATTCGTAGTTTCTGGAGTCTTGCT 60
QY 61 GGCCTATGATCAGCTGGTTACCCGTTGTTTACCCATGAAATGGCATCATCCGGC 120
Db 61 GGCCTATGATCAGCTGGTTACCCGTTGTTTACCCATGAAATGGCATCATCCGGC 120
QY 121 CCGAGCTCAACGTTAGACATTTATCAGTACGGTTCGGTAAAGCTGCGCTTCTGCAA 180
Db 121 CCGAGCTATGATCAGCTGGTTACCCGTTGTTTACCCATGAAATGGCATCATGCACTGCAA 180
QY 181 AGCGATCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCGCCAT 240
Db 181 AGCGATCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAACGGCGCCAT 240
QY 241 GTAGGCCAGGATGCGGATTAATAGTACTATTGAACTGACTGAAATGTTTCAAAATTAAT 300
Db 241 GTAGGCCAGGATGCGGATTAATAGTACTATTGAACTGACTGAAATGTTTCAAAATTAAT 300
QY 301 GCCACCATCGACAGTGGAAAGCTTAAAACTCCGATTTACTGTGCGCAATACGGCGGT 360
Db 301 GCCACCATCGACAGTGGAAAGCTTAAAACTCCGATTTACTGTGCGCAATACGGCGGT 360
QY 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATTCCAGCGTAAATGTCGTCAGATT 420
Db 361 AATAACGCGCGGCTGTTAATCAGACCGCATCTGATTCCAGCGTAAATGTCGTCAGATT 420
QY 421 GGTITGGCAACAGCCAGGCTAACGATATTA 456
Db 421 GGTITGGCAACAGCCAGGCTAACGATATTA 456

RESULT 7
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
XX AAC64628;
AC
XX 26-FEB-2001 (first entry)
XX
XX Agfa::PT#7 DNA sequence SEQ ID NO:23.
DE
XX Salmomella: agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
XX Salmomella enteritidis.
OS Escherichia coli.
OS Synthetic.
XX
XX WO20060102-A2.
PN
XX 12-OCT-2000.
PD
XX 05-APR-2000; 2000MO-CA000356.
PF
XX 05-APR-1999; 99US-0127888P.
PR
XX (UYVI-) UNIV VICTORIA.
PA
PI White AP, Doran JL, Collison SK, Kay WW;
P1 WPI: 2000-672631/65.
XX P-PSDB; AAB36352.
DR
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous recombination of a recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2.9e-106;

Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GCGCTATGATCAAGTGGTTCACCGTGTGTTACCCATGAATGACATGATCCGCG 120
DB 61 GCGCTATGATCAAGTGGTTCACCGTGTGTTACCCATGAATGACATGATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTGTGTTACCG 180
DB 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTGTGTTACCG 180
QY 181 AGCGATGCGCCGTAATTTGAAACGACCTTACCCAGACGGTTATGTTAAACGGCCGAT 240
DB 181 AGCGATGCGCCGTAATTTGAAACGACCTTACCCAGACGGTTATGTTAAACGGCCGAT 240
QY 241 GTAGGCCAGGGTGGGATATAGTATGAACTGATGAGTGAATGTTTCAAAATTAAT 300
DB 241 GTAGGCCAGGGTGGGATATAGTATGAACTGATGAGTGAATGTTTCAAAATTAAT 300
QY 301 GCCACATCGACGATGGAACGCTTAAAACTCCGATTTACTGCGGCAATACGGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTTAAAACTCCGATTTACTGCGGCAATACGGCGGT 360
QY 361 AATAAGCCCGGCTGTTAATCAGACCGCATGATTCAGCGCTATGTTGCTCAGGTT 420
DB 361 AATAAGCCCGGCTGTTAATCAGACCGCATGATTCAGCGCTATGTTGCTCAGGTT 420
QY 421 GGTTCGCAACGACGCGGCTTAAACAGTATTA 456
DB 421 GGTTCGCAACGACGCGGCTTAAACAGTATTA 456

```

RESULT 8
 AAC64622
 ID AAC64622 standard; DNA; 456 BP.

XX AAC64622;
 XX AAC64622;
 XX 26-FEB-2001 (first entry)
 XX

DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
 XX *Salmonella*: agfa; chromosomal gene replacement; fimbrial, epitope;
 KW vaccine; immune response; immunogen; de.
 XX
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 XX
 FN WO20060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CM000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UWVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WW;
 XX
 DR WPI: 2000-672631/65.
 XX
 DR P-PSDB; AAB56346.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC homologous recombination of a recombinant gene into the chromosome of the
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC the exemplification of the present invention

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2.9e-106;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GCGCTATGATCAAGTGGTTCACCGTGTGTTACCCATGAATGACATGATCCGCG 120
DB 61 GCGCTATGATCAAGTGGTTCACCGTGTGTTACCCATGAATGACATGATCCGCG 120
QY 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTGTGTTACCG 180
DB 121 CCGGACTCAACGTTGAGCATTTATAGTACCGTTCCGTTACCGTGTGTTACCG 180

```


PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUYI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay W;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB6347.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 CC
 SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;
 Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2.9e-106;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
 QY 1 ATGAACTTTTAAAGTGGAGCATTCGACGATCTGCTTCTGGCATGCTTGGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGACGATCTGCTTCTGGCATGCTTGGCT 60
 QY 61 GGGCTATGATGATGCTGCTTACCCGCTGTTTACCATTGAATGAGCATGATCGGCG 120
 DB 61 GGGCTGCTTCCAAATGGGGGGGGGGGGGATCATTAACGGCGGCAATATCTCGCG 120
 QY 121 CCGGACTCAACGTTGACATTTTACATGATCGGTCCGCTAACCGCTCGCTCTGCAA 180
 DB 121 CCGGACTCAACGTTGACATTTTACATGATCGGTCCGCTAACCGCTCGCTCTGCAA 180
 QY 181 AGCGATCCCGTAAATCTGAAGACCATTAACGAGGGGTTATGTAACGGCGCGAT 240
 DB 181 AGCGATCCCGTAAATCTGAAGACCATTAACGAGGGGTTATGTAACGGCGCGAT 240
 QY 241 GTAGGCGAGGGTGGGATTAATGATCTGTAAGTGAATGCTTTCAGAAATAT 300
 DB 241 GTAGGCGAGGGTGGGATTAATGATCTGTAAGTGAATGCTTTCAGAAATAT 300
 QY 301 GCACACATGACACAGTGGAAACGCTAAACCTCGATATTACTGTGGCCAAATAGCGGCT 360

DB 301 GCCACCATGACACAGTGGAAACGCTAAACCTCGATATTACTGTGGCCAAATATGATCAG 360
 QY 361 AATACCGCGCGCTGGTTAATCAGCCGATCTGATTCAGCGTAATGTCGTCAGGTT 420
 DB 361 CTGGTTACCGGTGTGTATCCCATGAATGCGACATGACAGCGTATGTCGTCAGGTT 420
 QY 421 GGTTTGGCAACAGCGGCTAACCGATTA 456
 DB 421 GGTTTGGCAACAGCGGCTAACCGATTA 456
 RESULT 11
 AAC64630
 ID AAC64630 standard; DNA; 456 BP.
 XX
 AC AAC64630;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#9 DNA sequence SEQ ID NO:27.
 XX
 KW *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 XX vaccine; immune response; immunogen; ds.
 OS *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000MO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UUYI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay W;
 XX WPI; 2000-672631/65.
 DR P-PSDB; AAB6347.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 XX protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAf) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP, 115 A, 116 C, 118 G, 107 T, 0 U, 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 3.3e-105;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATCCGCC 120
DB 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATCCGCC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTAAGCGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTAAGCGCTGCTCTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCGCGGTGGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCGCGGTGGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGATGGAAGCGTAAATCCGATATTATGTCGGCCAAATAGCGCGGT 360
DB 301 GCCACATCGACGATGGAAGCGTAAATCCGATATTATGTCGGCCAAATAGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTGTTGGCAACACGCGGCTTAACCGATTAATTA 456
DB 421 GGTGTTGGCAACACGCGGCTTAACCGATTAATTA 456

RESULT 12

AAC64627
ID AAC64627 standard; DNA; 456 BP.

AC AAC64627;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#6 DNA sequence SEQ ID NO:21.

KV Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

KW vaccine; immune response; immunogen; ds.

OS Salmonella enteritidis.

OS Becherichia coli.

OS Synthetic.

PN WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay MW;

XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (1) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SFP1/TF) nucleation depended
CC assembly system of strains of Salmonella, Escherichia coli and
CC Enterobacteriaceae for the production of fimbriae comprising recombinant
CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant Agfa
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmonella, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP, 112 A, 113 C, 125 G, 106 T, 0 U, 0 Other;

Query Match 76.5%; Score 348.8; DB 3; Length 456;
Best Local Similarity 85.3%; Pred. No. 3.3e-105;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTAGTTCTGGAGTCTTGCT 60
QY 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATCCGCC 120
DB 61 GCGCTCTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCAGCATCCGCC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTAAGCGCTGCTCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTAAGCGCTGCTCTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCGCAT 240
QY 241 GTAGCCGCGGTGGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGCCGCGGTGGCGATTAATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGATGGAAGCGTAAATCCGATATTATGTCGGCCAAATAGCGCGGT 360
DB 301 GCCACATCGACGATGGAAGCGTAAATCCGATATTATGTCGGCCAAATAGCGCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTGTTGGCAACACGCGGCTTAACCGATTAATTA 456
DB 421 GGTGTTGGCAACACGCGGCTTAACCGATTAATTA 456

PA (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
PA (KING/) KING J.
PI Kay WW, Collinson SK, Clouthier SC, Doran JL;
XX WPI; 1994-358275/44.
DR P-PSDB; AAR62761.
XX
PT Eliciting an immune response to Salmonella - using attenuated Salmonella
PT strains, vector constructs, or compans. contg. fimbrial type proteins.
XX
XX Disclosure; Fig 7A; 95pp; English.
XX
XX The DNA encodes the Salmonella enteritidis 27655-3b *trpH* mutant strain
CC *agfA* gene cloned into pUC19. The DNA and isolated proteins are used in
CC genetic immunization and vaccine compositions, respectively, to elicit an
CC immune response to Salmonella in animals (e.g. food producing animals)
CC and humans. (Updated on 25-MAR-2003 to correct PN field.) (Updated on 27-
CC AUG-2003 to correct OS field.)
XX
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
Query Match 62.3%; Score 284.2; DB 2; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.4e-84;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 110 ATGATCCGCGCCGAGCTCAACGTGAGCATTTATCAGTTCGCTACCGTGGC 169
DB 47 ATAGTTCCGCGCCGAGCTCAACGTGAGCATTTATCAGTTCGCTACCGTGGC 106
QY 170 TTGCTCTGCAAAAGCATGCCGCTTAATCTGAAACGACATTACCGAGCGGTTATGTA 229
DB 107 TTGCTCTGCAAAAGCATGCCGCTTAATCTGAAACGACATTACCGAGCGGTTATGTA 166
QY 230 ACGGCGCGGATGTAGGCGGAGGTGGGATTAATGACTATTGAACTGACTGAGATGGT 289
DB 167 ACGGCGCGGATGTAGGCGGAGGTGGGATTAATGACTATTGAACTGACTGAGATGGT 226
QY 290 TCAGAAATTAATGCCCATTCGACCGAGTGAACGCTTAATACTCGATTTACTGCGGCC 349
DB 227 TCAGAAATTAATGCCCATTCGACCGAGTGAACGCTTAATACTCGATTTACTGCGGCC 286
QY 350 AATACGGCGGTAATTAACGCCGCGGTGTTAATCAGACCGCATCTGATTC 398
DB 287 AATACGGCGGTAATTAACGCCGCGGTGTTAATCAGACCGCATCTGATTC 335
RESULT 15
AAT74141
ID AAT74141 standard; DNA; 361 BP.
XX
AC AAT74141;
XX
DT 25-MAR-2003 (revised)
DT 30-SEP-1997 (first entry)
XX
DB Salmonella enteritidis 27655-3b *trpH* mutant *agfA* gene fragment.
XX
KM Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; da.
XX
OS Salmonella enteritidis.
XX
FH Key Location/Qualifiers
FT CDS 1..360
FT /*tag= a
FT /*label= *agfA*_gene_fragment
FT 16..60
FT /*tag= b
FT /*label= Primer_TAF5
FT 52..69
FT /*tag= c
FT /*label= Primer_TAF3
FT primer_bind complement(103..128)

FT /*tag= d
FT /*label= Primer_TAF6
FT primer_bind complement(294..312)
FT /*tag= e
FT /*label= Primer_TAF4
XX
XX US5635617-A.
XX
XX 03-JUN-1997.
XX
XX 26-APR-1994; 94US-00233788.
XX
XX 26-APR-1993; 93US-00054452.
XX
XX (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX Collinson SK, Kay WW, Doran JL;
XX WPI; 1997-309886/28.
XX P-PSDB; AAM23569.
XX
XX Isolated Salmonella gene *agfA* - used for diagnosis of Salmonella or
XX enteropathogenic bacteria of the Enterobacteria family.
XX
XX Claim 1; Col 107-110; 85pp; English.
XX
XX The present sequence represents an isolated *agfA* gene fragment derived
XX from Salmonella enteritidis 27655-3b *trpH* mutant strain. The nucleic
XX acid can be used to provide diagnostic assays for Salmonella and/or
XX enteropathogenic bacteria of the family Enterobacteria. It can also be
XX used to provide proteins and antibodies which can be used for assays. The
XX nucleic acid sequence can be used to provide probes or primers which can
XX specifically hybridise to nucleic acid molecules from greater than 99% of
XX Salmonella strains that are pathogenic to warm-blooded animals relative
XX to nucleic acid molecules from virtually all other microbial organisms.
XX (Updated on 25-MAR-2003 to correct PF field.)
SQ Sequence 361 BP; 94 A; 93 C; 94 G; 80 T; 0 U; 0 Other;
Query Match 62.3%; Score 284.2; DB 2; Length 361;
Best Local Similarity 99.0%; Pred. No. 8.4e-84;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 110 ATGATCCGCGCCGAGCTCAACGTGAGCATTTATCAGTTCGCTACCGTGGC 169
DB 47 ATAGTTCCGCGCCGAGCTCAACGTGAGCATTTATCAGTTCGCTACCGTGGC 106
QY 170 TTGCTCTGCAAAAGCATGCCGCTTAATCTGAAACGACATTACCGAGCGGTTATGTA 229
DB 107 TTGCTCTGCAAAAGCATGCCGCTTAATCTGAAACGACATTACCGAGCGGTTATGTA 166
QY 230 ACGGCGCGGATGTAGGCGGAGGTGGGATTAATGACTATTGAACTGACTGAGATGGT 289
DB 167 ACGGCGCGGATGTAGGCGGAGGTGGGATTAATGACTATTGAACTGACTGAGATGGT 226
QY 290 TCAGAAATTAATGCCCATTCGACCGAGTGAACGCTTAATACTCGATTTACTGCGGCC 349
DB 227 TCAGAAATTAATGCCCATTCGACCGAGTGAACGCTTAATACTCGATTTACTGCGGCC 286
QY 350 AATACGGCGGTAATTAACGCCGCGGTGTTAATCAGACCGCATCTGATTC 398
DB 287 AATACGGCGGTAATTAACGCCGCGGTGTTAATCAGACCGCATCTGATTC 335
Search completed: March 15, 2004, 17:51:59
Job time : 246.684 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds
(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....ccacgcctaacagatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA:*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PTCUTS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	403.2	88.4	456	1 US-08-233-788A-58	Sequence 58, Appl
2	284.2	62.3	361	1 US-08-233-788A-56	Sequence 56, Appl
3	33.4	7.3	7766	4 US-09-125-619-3	Sequence 3, Appl
4	31.2	6.8	1491	6 5486473-3	Parent No. 5486473
5	31.2	6.8	2436	4 US-09-540-236-492	Sequence 492, App
6	31.2	6.8	65792	4 US-09-596-002-31	Sequence 31, Appl
7	30.8	6.8	2547	3 US-08-508-761B-1	Sequence 1, Appl
8	30.8	6.8	1664976	4 US-08-916-421B-1	Sequence 1, Appl
9	30.2	6.6	1155	4 US-09-328-352-1775	Sequence 1775, Ap
10	29.6	6.5	972	4 US-09-328-352-2055	Sequence 2055, Ap
11	29.4	6.4	1939	4 US-08-961-527-310	Sequence 310, App
12	29.4	6.4	3494	3 US-09-133-802-200	Sequence 200, App
13	29.4	6.4	3494	4 US-09-659-786-200	Sequence 200, App
14	29.4	6.4	3494	4 US-09-023-655-1006	Sequence 1006, Ap
15	29.2	6.4	2052	1 US-08-443-104-5	Sequence 5, Appl
16	29.2	6.4	2052	1 US-08-238-130-6	Sequence 6, Appl
17	29.2	6.4	2052	1 US-08-442-859-5	Sequence 5, Appl
18	29.2	6.4	2052	2 US-08-396-489-5	Sequence 5, Appl
19	29.2	6.4	2052	2 US-08-894-772-1	Sequence 1, Appl
20	29.2	6.4	2052	2 US-09-207-844-1	Sequence 1, Appl
21	29.2	6.4	2052	4 US-09-253-509-1	Sequence 1, Appl
22	29.2	6.4	2052	5 PCT-US95-05534-5	Sequence 5, Appl
23	29	6.4	13794	4 US-08-956-171E-54	Sequence 54, Appl
24	29	6.4	1830121	4 US-09-557-884-1	Sequence 1, Appl
25	29	6.4	1830121	4 US-09-643-990A-1	Sequence 1, Appl
26	28.8	6.3	720	4 US-09-328-352-354	Sequence 354, App
27	28.8	6.3	1662	1 US-08-565-386-2	Sequence 2, Appl

28	28.8	6.3	1836	4 US-09-252-991A-15831	Sequence 15831, A
29	28.8	6.3	2055	4 US-09-252-991A-15771	Sequence 15771, A
30	28.8	6.3	2106	4 US-09-252-991A-15891	Sequence 15891, A
31	28.8	6.3	4529	1 US-08-565-386-1	Sequence 1, Appl
32	28.8	6.3	8906	4 US-09-027-169-5	Sequence 5, Appl
33	28.4	6.2	1520	4 US-09-484-970B-30	Sequence 30, Appl
34	28.2	6.2	2116	3 US-09-000-041A-1	Sequence 1, Appl
35	28.2	6.2	2426	3 US-09-028-327-2	Sequence 2, Appl
36	28.2	6.2	2426	4 US-09-571-078A-2	Sequence 2, Appl
37	28.2	6.2	2812	1 US-08-920-812-16	Sequence 16, Appl
38	28.2	6.2	2812	1 US-08-920-827-16	Sequence 16, Appl
39	28.2	6.2	2812	1 US-08-921-177-16	Sequence 16, Appl
40	28.2	6.2	2812	1 US-08-362-577C-16	Sequence 16, Appl
41	28.2	6.2	2812	2 US-08-920-828-16	Sequence 16, Appl
42	28.2	6.2	4973	3 US-09-381-862-6	Sequence 6, Appl
43	28.2	6.2	8078	4 US-09-702-251-3	Sequence 3, Appl
44	28	6.1	906	4 US-09-134-001C-383	Sequence 383, App
45	28	6.1	1665	4 US-09-540-236-1858	Sequence 1858, App

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 372836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 88.4%; Score 403.2; DB 1; Length 456;
Best Local Similarity 92.8%; Pred. No. 5e-135;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

RESULT 2
 US-08-233-788A-56
 Sequence 56, Application US/08233788A
 Patent No. 5635617
 GENERAL INFORMATION:
 APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Collinson, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Avenue, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043.403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEDANBERRY
 INFORMATION FOR SEQ ID NO: 56:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

Query Match	62.3%	Score 284.2	DB 1	Length 361.1
Best Local Similarity	99.0%	Pred. No. 2.9e-92		
Matches 286; Conservative	0	Mismatches 3	Indels 0	

Qy 110 ATGCATCCGGCCCGGACTCAACGTTGACATTATTCAGTACGGTCCGCTAACGCTGGC 169
Db 47 ATAGTTCGGGCCCGGAGCTACAGTTGAGACTTTATCAGTACGGTTCGCTAACGCTGGC 106
Qy 170 TTGCTCTGCAAAAGCGATGCGCGCTAAATCTGAAAGACCATTAACCGACCGGTTATGGTA 229
Db 107 TTGCTCTGCAAAAGCGATGCGCGCTAAATCTGAAAGACCATTAACCGACCGGTTATGGTA 166
Qy 230 ACGGCGCGGATGTAGAGCCAGAGGTCGGATTAATGTACTTGAACGTGACTCAGAAATGTT 289
Db 167 ACGGCGCGGATGTAGAGCCAGAGGTCGGATTAATGTACTTGAACGTGACTCAGAAATGTT 226
Qy 290 TCGAATAATTAATCCACCATGACCCAGTGGAAAGCCTTAAAACTCGGATTAATCTGTGGCC 349
Db 227 TCGAATAATTAATCCACCATGACCCAGTGGAAAGCCTTAAAACTCGGATTAATCTGTGGCC 286
Qy 350 AATACGGCGGTTAATTAACGCGCGCTGTTAATTCACACCGCATTCGATTC 398
Db 287 AATACGGCGGTTAATTAACGCGCGCTGTTAATTCACACCGCATTCGATTC 335

```

RESULT 3
US-09-125-619-3
: Sequence 3, Application US/09125619
: Patent No. 643716
: GENERAL INFORMATION:
: APPLICANT: NORRIS, STEVEN J.
: APPLICANT: JING-REN, ZHANG
: APPLICANT: HARDAM, JOHN M.
: APPLICANT: HOWELL, JERRILYN K.
: APPLICANT: BARBOUR, ALAN G.
: TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELLIA
: FILE REFERENCE: UTS# 234
: CURRENT APPLICATION NUMBER: US/09/125,619
: NUMBER OF SEQ ID NOS: 48
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 7766
: TYPE: DNA
: ORGANISM: Borrelia burgdorferi
: FEATURES:
: NAME/KEY: modified_base
: LOCATION: (127)
: OTHER INFORMATION: R = A OR G
US-09-125-619-3

```

	Query Match	Best Local Similarity	7.38%	Score 33.4	DB 4	Length 7766
Matches	100	Conservative	0	Mismatches	111	Indels
						Gaps
QY	157	GCTAACGCTGC	CGCTTGTCTGC	CAAAAGGAGCCCGCTAATCTGAACGACCATTA	CCCGAG	216
Db	4432	GTTACGGCTGCGCGCTGCTGTGACAGAGATGAGAGAGAGCTGCAAGCTCAACAAATCG	4491			
QY	217	AGGGGTTATGTGTAACGCGCCGATGTAGGCGAGTGTGCGGATTAATGTCTATTGAAC	CTG	276		
Db	4492	ATTGCTGCTGCTATTGTGGAAAGGTAAATGAGAGATGTGGCGGATTTTGGTAGGATTAAGATG	4551			


```

1 LOCATION: (28257) .. (28258)
2 OTHER INFORMATION: n equals a, t, c, or g
3 NAME/KEY: misc.feature
4 LOCATION: (84773) .. (84773)
5 OTHER INFORMATION: n equals a, t, c, or g
6 NAME/KEY: misc.feature
7 LOCATION: (84808) .. (84808)
8 OTHER INFORMATION: n equals a, t, c, or g
9 NAME/KEY: misc.feature
10 LOCATION: (84812) .. (84812)
11 OTHER INFORMATION: n equals a, t, c, or g
12 NAME/KEY: misc.feature
13 LOCATION: (98120) .. (98120)
14 OTHER INFORMATION: n equals a, t, c, or g
15 NAME/KEY: misc.feature
16 LOCATION: (98159) .. (98159)
17 OTHER INFORMATION: n equals a, t, c, or g
18 NAME/KEY: misc.feature
19 LOCATION: (98239) .. (98239)
20 OTHER INFORMATION: n equals a, t, c, or g
21 NAME/KEY: misc.feature
22 LOCATION: (98266) .. (98266)
23 OTHER INFORMATION: n equals a, t, c, or g
24 NAME/KEY: misc.feature
25 LOCATION: (98343) .. (98343)
26 OTHER INFORMATION: n equals a, t, c, or g
27 NAME/KEY: misc.feature
28 LOCATION: (103998) .. (103998)
29 OTHER INFORMATION: n equals a, t, c, or g
30 NAME/KEY: misc.feature
31 LOCATION: (148948) .. (148948)
32 OTHER INFORMATION: n equals a, t, c, or g
33 NAME/KEY: misc.feature
34 LOCATION: (163385) .. (163385)
35 OTHER INFORMATION: n equals a, t, c, or g
36 NAME/KEY: misc.feature
37 LOCATION: (191989) .. (191989)
38 OTHER INFORMATION: n equals a, t, c, or g
39 NAME/KEY: misc.feature
40 LOCATION: (191995) .. (191995)
41 OTHER INFORMATION: n equals a, t, c, or g
42 NAME/KEY: misc.feature
43 LOCATION: (231980) .. (231980)
44 OTHER INFORMATION: n equals a, t, c, or g
45 NAME/KEY: misc.feature
46 LOCATION: (234187) .. (234187)
47 OTHER INFORMATION: n equals a, t, c, or g
48 NAME/KEY: misc.feature
49 LOCATION: (234220) .. (234220)
50 OTHER INFORMATION: n equals a, t, c, or g
51 NAME/KEY: misc.feature
52 LOCATION: (234814) .. (234814)
53 OTHER INFORMATION: n equals a, t, c, or g
54 NAME/KEY: misc.feature
55 LOCATION: (309398) .. (309398)
56 OTHER INFORMATION: n equals a, t, c, or g
57 NAME/KEY: misc.feature
58 LOCATION: (309418) .. (309418)
59 OTHER INFORMATION: n equals a, t, c, or g
60 NAME/KEY: misc.feature
61 LOCATION: (312837) .. (312837)
62 OTHER INFORMATION: n equals a, t, c, or g
63 NAME/KEY: misc.feature
64 LOCATION: (312993) .. (312993)
65 OTHER INFORMATION: n equals a, t, c, or g
66 NAME/KEY: misc.feature
67 LOCATION: (319326) .. (319326)
68 OTHER INFORMATION: n equals a, t, c, or g
69 NAME/KEY: misc.feature
70 LOCATION: (559167) .. (559167)
71 OTHER INFORMATION: n equals a, t, c, or g
72 NAME/KEY: misc.feature
73 LOCATION: (559241) .. (559241)

```

```

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (600592)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g

```

```

NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

Query Match 6.8%; Score 30.8; DB 4; Length 1664976;

Best Local Similarity 51.4%; Pred. No. 60;

Matches 71; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

Qy 288 TTTCAGAAATATACCCACCAATGACCGAGTGAAGCTTAACCTCCATATTACTGCGG 347
Db 360315 TTTTATTAATGTTGTCATGATGTCGTAACGAAATATACGTTTATTTATCTCAT 360374
Qy 348 CCAATACGCGCGGTATATACCGCGCTGTATATCAGACCCGATCTGATTCAGCGTAAT 407
Db 360375 CAACGGTGAACAAAGAAACAAAGCTTTATGATATGAACTTTTAAATCATTTAAG 360434
Qy 408 GGTGCGTCAGGTGCTTT 425
Db 360435 GATGCTTAGGCTTTTAT 360452

```

RESULT 9

US-09-328-352-1775

Sequence 1775, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE REFERENCE: GTC99-03PA

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 1775

LENGTH: 1155

TYPE: DNA

ORGANISM: Acinetobacter baumannii

US-09-328-352-1775

Query Match 6.8%; Score 30.2; DB 4; Length 1155;

Best Local Similarity 51.9%; Pred. No. 1.5;

Matches 68; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

Qy 79 GTTACCGGTGTTTACCCATGAATGACATGATCCGCGCGACTCAAGTTGAGC 138
Db 799 GATATTGAGCAATATATCAGCAATGCTTATGATATGCTATGACCAAACTTGGTA 858
Qy 139 ATTATACGATACGCGTTCGCTAACGCTGCGCTTCTGCAAAAGCATCCGTAATCT 198
Db 859 ATTATCAAAAGTCTCAAGGAAACTCAATCTTCACTATTTGTTGACCAAAATCT 918
Qy 199 GAAACGACCAT 209
Db 919 AAACGACCAT 929

```

RESULT 10

US-09-328-352-2055/C

Sequence 2055, Application US/09328352

Patent No. 6562958

GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER

FILE REFERENCE: GTC99-03PA

CURRENT FILING DATE: 1999-06-04

NUMBER OF SEQ ID NOS: 8252

SEQ ID NO 2055

LENGTH: 972
 TYPE: DNA
 ORGANISM: Acinetobacter baumannii
 US-09-328-352-2055

Query Match
 Best Local Similarity 6.5%; Score 29.6; DB 4; Length 972;
 Matches 47; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 265 ACTATTGAAGTCTGAGATGGTTTTCAGAAATTAATGCCACCATGACGACGATGAGAGCGT 324
 DB 558 ATTAGTGAAGTAAATTAATTAATGACCATATAATAGCCCACTGAGCTTATGATGATGACT 499
 QY 325 AAAAATCCGATATTA 340
 DB 498 CAATACCTCTAGATTA 483

RESULT 11

US-08-961-527-310
 Sequence 310 Application US/08961527
 Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunech
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,527
 FILING DATE:

CLASSIFICATION:

CLASSIFICATION: 424
 PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO:

SEQUENCE CHARACTERISTICS:
 LENGTH: 1939 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-527-310

Query Match

Best Local Similarity 6.4%; Score 29.4; DB 4; Length 1939;
 Matches 54; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 235 GCCGATGTAGCCAGGGTGGGATATAGTACTATTGAATGCTGACGAGATGTTTGA 294
 DB 618 GCCATGTATACAGGTTAGGGAAGAAAGACTTACTGATTTCTTAAGAGTACACCACT 677
 QY 295 AATTAAGCCACATGACGACGATGAGTAAAGGCTAATAA 329
 DB 678 AAGGTCTCTACTGATAGTAAAGTACTGCTGAAA 712

RESULT 12

US-09-139-802-200/c
 Sequence 200 Application US/09139802
 Patent No. 6180084

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki
 TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing
 TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using
 FILE REFERENCE: P-1J 3203

CURRENT APPLICATION NUMBER:

CURRENT FILING DATE: US/09/139,802
 EARLIER APPLICATION NUMBER: 1998-08-25

EARLIER FILING DATE: 1997-09-10
 EARLIER APPLICATION NUMBER: 08/926,914

EARLIER FILING DATE: 1996-09-10
 NUMBER OF SEQ ID NOS: 226

SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 200

LENGTH: 3494

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS
 LOCATION: (121)..(3024)

US-09-139-802-200

Query Match

Best Local Similarity 6.4%; Score 29.4; DB 3; Length 3494;
 Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGTTATGTATGTAACGGCGCGATGATGAGCCAGGGTGGATTAATAGTATTG 271
 DB 2531 CCTGGGAGATGAGGTTGCACTGATGACGATGACCGGAGTTGGGATGATCGGTTATTA 2472
 QY 272 AACTGACTCAGAAATGCTTTCAGAAATATGACCAATGACCAAGTGAAGGCTAATAAACT 331
 DB 2471 TTGGGTTTCCATCCACTCTTGAAAAGCCAGAGACCATCTCCACACTCTGAAGT 2412
 QY 332 CCGATATTACTGTGCGCAATACGGCGGTAATTAACCGCGCTGGTATATACGACGAT 391
 DB 2411 CCGTTGAGAGAGGGGCGTGTGATGATTAACCTGCTGTACTGTGATGATGATGATGAT 2352
 QY 392 CTGATTCAGCGTAATGATGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 450
 DB 2251 GGGATCTCCCTCCAGTTGTTGATTAATTTCTGAAGTGAATGAAGAGGGGTGTGACCTG 2293

RESULT 13

US-09-659-786-200/c
 Sequence 200 Application US/09659786
 Patent No. 6491894

GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki
 TITLE OF INVENTION: NGR Receptor and Methods of Identifying Tumor Homing
 TITLE OF INVENTION: Molecules That Home to Angiogenic Vasculature Using

FILE REFERENCE: P-1J 3203

CURRENT APPLICATION NUMBER:

CURRENT FILING DATE: US/09/659,786
 PRIOR APPLICATION NUMBER: 2000-09-11

PRIOR FILING DATE: 1997-09-10
 PRIOR APPLICATION NUMBER: 08/926,914

PRIOR FILING DATE: 1996-09-10
 NUMBER OF SEQ ID NOS: 226

SOFTWARE: Patentin Ver. 2.0
 SEQ ID NO 200

LENGTH: 3494

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS
LOCATION: (121) (3024)
US-09-659-786-200

Query Match 6.4%; Score 29.4; DB 4; Length 3494;
Best Local Similarity 45.2%; Pred. No. 5.6;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTAAGTAAAGCCCGCATGTAAGCCAGGTCGGATAATAGTACTATG 271
2531 CCCGCGGAGTACGCTTACAGTACGCTGACCGCAGGTTGGGTGAGTGGATTATTA 2472
272 AACTGACTCAGATGCTTTCAGAAATATATGCCACCATGACCACTGAAACCTTAAACT 331
2471 TTGGGGTTCTCATCTCATCTGTTGAAAAGCCAGACATCTCTCATCTGAACT 2412
332 CCGATATTACTGTGCGCAATACGCGGTTAATACCGCGCTGTTATACAGCCGAT 391
2411 CCGTTGAGCAGCGGCTGCTGATGACATTAACCTGCTGTAAGTCACTGAGTTTCT 2352
392 CTGATTCAGCGTATGTTGCGTCAGGTTGTTGGCAACAACGCCACGCTAACCG 450
2351 GGGATCTCCCTCCAGTTGTTGTTATTTCTGAAATGAAGAGGGGTGTGACCTG 2293

RESULT 14

US-09-023-655-1006/c
Sequence 1006, Application US/09023655

PATENT No. 6607879
GENERAL INFORMATION:
APPLICANT: Cooke, Benjamin G.
APPLICANT: Susan G. Stuart
TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF BLOOD CELL GENE
NUMBER OF SEQUENCES: 1508
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023.655
FILING DATE: HEREMITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Zeller, Karen J.
REGISTRATION NUMBER: 37,071
REFERENCE/DOCKET NUMBER: PA-0001 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 855-0555
TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 1006:
SEQUENCE CHARACTERISTICS:
LENGTH: 3494 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GENBANK
CLONE: g178535
US-09-023-655-1006

Query Match 6.4%; Score 29.4; DB 4; Length 3494;
Best Local Similarity 45.2%; Pred. No. 5.6;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTAAGTAAAGCCCGCATGTAAGCCAGGTCGGATAATAGTACTATG 271
2531 CCCGCGGAGTACGCTTACAGTACGCTGACCGCAGGTTGGGTGAGTGGATTATTA 2472
272 AACTGACTCAGATGCTTTCAGAAATATATGCCACCATGACCACTGAAACCTTAAACT 331
2471 TTGGGGTTCTCATCTCATCTGTTGAAAAGCCAGACATCTCTCATCTGAACT 2412
332 CCGATATTACTGTGCGCAATACGCGGTTAATACCGCGCTGTTATACAGCCGAT 391
2411 CCGTTGAGCAGCGGCTGCTGATGACATTAACCTGCTGTAAGTCACTGAGTTTCT 2352
392 CTGATTCAGCGTATGTTGCGTCAGGTTGTTGGCAACAACGCCACGCTAACCG 450
2351 GGGATCTCCCTCCAGTTGTTGTTATTTCTGAAATGAAGAGGGGTGTGACCTG 2293

RESULT 15

US-08-443-104-5/c
Sequence 5, Application US/08443104

PATENT No. 5691162
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglesand, Claus
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5691162 disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443.104
FILING DATE: 17-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398.489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr. Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-443-104-5

Query Match 6.4%; Score 29.2; DB 1; Length 2052;
Best Local Similarity 49.4%; Pred. No. 4.9;
Matches 76; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

OY 300 TGGACCATGACAGTGAACGCTAAAACTCGATATTACTGTGGCCCAATACGGCG 359
 Db 563 TGTAGCCTCAACGGGAGAGAAACATCAACGGTGTCTTGAGAGCGTCGACGGGGT 504
 OY 360 TAATACGCCCGCGCTGTTATCATCAGACCGCATCTGATTCAGGTAATGTGGTCAGGT 419
 Db 503 CTTTCTGTCACCGCTTGTAAAGAGACCGGGAATCTTGCCCTGTGAAGACTGTTCGGT 444
 OY 420 TGGTTTGGCAACAGCCACGGCTAACCAATAT 453
 Db 443 AGAGAGAGACCTCGCGCTACGCCCAATCTGTAT 410

Search completed: March 16, 2004, 04:37:30
 Job time : 52.6647 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atggaacttcaaaagtcgc.....ccacgcctaccagtattaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35.6	7.8	1965	9 US-09-815-242-6264	Sequence 6264, Ap
2	35.6	7.8	1965	12 US-10-282-122A-20332	Sequence 20332, A
3	34	7.5	2583	12 US-10-282-122A-19592	Sequence 19592, A
4	34	7.5	2886	9 US-09-801-368-131	Sequence 131, App
5	33.6	7.4	549	10 US-09-991-936-1800	Sequence 1800, Ap
6	33.4	7.3	7766	12 US-10-222-566-3	Sequence 3, Appl1
7	33.4	7.3	7766	14 US-10-222-162-3	Sequence 3, Appl1
8	33.4	7.3	1059	9 US-09-738-626-482	Sequence 482, App
9	32.8	7.2	1113	9 US-09-738-626-481	Sequence 481, App
10	32.8	7.2	3309400	9 US-09-738-626-1	Sequence 1, Appl1
11	32.8	7.1	568	14 US-10-333-631-1	Sequence 4, Appl1
12	32.2	7.1	972	14 US-10-333-631-1	Sequence 4, Appl1
13	32.2	7.1	1083	15 US-10-369-493-33133	Sequence 33133, A
14	32.2	7.1	4830	12 US-10-282-122A-27869	Sequence 27869, A
15	32.2	7.1	4830	12 US-10-282-122A-27869	Sequence 27869, A

16	31.8	7.0	930	9 US-09-815-242-6028	Sequence 6028, Ap
17	31.8	7.0	930	12 US-10-282-122A-20335	Sequence 20335, A
18	31.8	7.0	930	15 US-10-369-493-24480	Sequence 24480, A
19	31.2	6.8	669	14 US-10-156-761-687	Sequence 687, App
20	31.2	6.8	2394	12 US-10-282-122A-26837	Sequence 26837, A
21	31.2	6.8	9025608	14 US-10-156-761-1	Sequence 1, Appl1
22	31	6.8	525	10 US-09-918-995-32144	Sequence 32144, A
23	31	6.8	616	12 US-10-424-599-88945	Sequence 88945, A
24	31	6.8	2751	15 US-10-398-221-3563	Sequence 3563, App
25	30.8	6.8	715	12 US-10-424-599-10132	Sequence 10132, A
26	30.8	6.8	1575	10 US-09-873-367C-255	Sequence 255, App
27	30.8	6.8	1665	15 US-10-369-493-43454	Sequence 43454, A
28	30.8	6.8	2050	14 US-10-086-510-1	Sequence 1, Appl1
29	30.6	6.7	1668	9 US-10-355-956-3	Sequence 3, Appl1
30	30.6	6.7	2301	9 US-09-815-242-6844	Sequence 4844, Ap
31	30.6	6.7	2301	9 US-09-815-242-5042	Sequence 9042, Ap
32	30.4	6.7	392	12 US-10-282-122A-25231	Sequence 25231, A
33	30.4	6.7	25220	11 US-09-984-429-364	Sequence 364, App
34	30.2	6.6	379	12 US-10-424-599-102524	Sequence 102524, A
35	30.2	6.6	3203	12 US-10-282-122A-24039	Sequence 24039, A
36	30.2	6.6	3222	12 US-10-282-122A-6780	Sequence 6780, App
37	30.2	6.6	3222	14 US-10-210-115-19	Sequence 19, Appl1
38	30.2	6.6	3222	15 US-10-369-493-24379	Sequence 24379, A
39	30.2	6.6	1601042	15 US-10-027-632-59064	Sequence 59064, A
40	30	6.6	414	14 US-10-156-761-2947	Sequence 2947, Ap
41	30	6.6	1074	12 US-10-282-122A-30822	Sequence 30822, A
42	29.8	6.5	436	12 US-10-424-599-73842	Sequence 73842, A
43	29.8	6.5	815	12 US-10-424-599-72759	Sequence 72759, A
44	29.8	6.5	1394	14 US-10-283-423-13	Sequence 13, Appl1
45	29.8	6.5	1394	14 US-10-213-821-13	Sequence 13, Appl1

ALIGNMENTS

RESULT 1
US-09-815-242-6264
Sequence 6264, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zykkind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Treawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: PacSeq for Windows Version 4.0
SEQ ID NO 6264
LENGTH: 1965
TYPE: DNA
ORGANISM: Escherichia coli

FEATURE:
NAME/KEY: CDS
LOCATION: (1)... (1965)
US-09-815-242-6264

Query Match
Best Local Similarity 7.8%; Score 35.6; DB 9; Length 1965;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 41 TTCTGGCACTGCTCTGGCTGGCTGTATGATCACTGGTTACCGGTGTTTACCATG 100
DB 350 TTCTGACAGTCTCTGCTCCAGGCGTAGTGATGATGTGTCACCGAATGTACCGCTTG 409
QY 101 AATGGCAATGATCCGCGCCGAGCTCAAGCTTGAGC 138
DB 410 AGAAGTTCCTGCTGCTGACCTGGCCCCCTGCTCCG 447

RESULT 2

US-10-282-122A-20532
Sequence 20532, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 20532
LENGTH: 1965
TYPE: DNA
ORGANISM: Escherichia coli
US-10-282-122A-20532

Query Match
Best Local Similarity 7.8%; Score 35.6; DB 12; Length 1965;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 41 TTCTGGCACTGCTCTGGCTGGCTGTATGATCACTGGTTACCGGTGTTTACCATG 100

DB 350 TTCTGACAGTCTCTGCTCCAGGCGTAGTGATGATGTGTCACCGAATGTACCGCTTG 409
QY 101 AATGGCAATGATCCGCGCCGAGCTCAAGCTTGAGC 138
DB 410 AGAAGTTCCTGCTGCTGACCTGGCCCCCTGCTCCG 447

RESULT 3

US-10-282-122A-19592
Sequence 19592, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.

FILE REFERENCE: ELITRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 19592
LENGTH: 2583
TYPE: DNA
ORGANISM: Enterobacter cloacae
US-10-282-122A-19592

Query Match
Best Local Similarity 7.5%; Score 34; DB 12; Length 2583;
Matches 91; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 68 ATGATCACTGCTGATCCGCTGTTGTTACCATGAATGACATGATCCGCCGACT 127
DB 1 ATGCTCAACCGAATTATCGTTGATCGTCCGCTGAGAGGCTTCTGTTGAACTGACG 60
QY 128 CAAGTTGAGATTATCACTGATCGTTCCGCTAACCGCTGCTGCTGCAAAACGATG 187
DB 61 GCGCGCAGGCGATGCTCGAGTGTGCGCTGACGCTGACCGTCTGCGACGACGCG 120
QY 188 CCGTAAATCTGAAGACCATTAACCCAGAGGGTTATGATGAGCGCCGATGAGCC 247
DB 121 CGCATGACCGCAGTAGTGTGCTGGCCAGCCGCTGACGATCCATCCGACGACGAGC 180

RESULT 7

US-10-222-162-3
; Sequence 3, Application US/10222162
; Publication No. US20030060618A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: US/234USD4
; CURRENT APPLICATION NUMBER: US/10/222,162
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (127)
; OTHER INFORMATION: R = A OR G
US-10-222-162-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3; Mismatches 111; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCTGCTCTCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 216
DB 4432 GTTACGCTGCTGCTGCTCTGCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 4491
QY 217 AGCGTTATGTATGTAACGCGCGATGTAGCGCCAGGCGGTGGATTAATAGTATTGAAC 276
DB 4492 ATTGCTGCTGCTGCTGCTGCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 4551
QY 277 ACTCAGATGTTTCAAGAAATATATCCATCCATCCAGTGAAGCGTAAATACTCCGAT 336
DB 4552 AAGAAAGATGATCAGATTGCTGCTCTATGCTTTGAGGGGATGCTTAAGATGGAAG 4611
QY 337 ATTACTGTGCGCAATAGCGCGTAAATACG 367
DB 4612 TTGCTGTGAAGATGATGATGATGGAAG 4642

RESULT 8

US-10-143-024-3
; Sequence 3, Application US/10143024
; Publication No. US20030092903A1
; GENERAL INFORMATION:
; APPLICANT: NORRIS, STEVEN J.
; APPLICANT: JING-REN, ZHANG
; APPLICANT: HARDHAM, JOHN M.
; APPLICANT: HOWELL, JERRILYN K.
; APPLICANT: BARBOUR, ALAN G.
; APPLICANT: WEINSTOCK, GEORGE M.
; TITLE OF INVENTION: VMP-LIKE SEQUENCES OF PATHOGENIC BORRELIA
; FILE REFERENCE: US/234USD4
; CURRENT APPLICATION NUMBER: US/10/143,024
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: 09/125,619
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 7766
; TYPE: DNA
; ORGANISM: Borrelia burgdorferi

FEATURE:
NAME/KEY: modified_base
LOCATION: (127)
OTHER INFORMATION: R = A OR G
US-10-143-024-3

Query Match 7.3%; Score 33.4; DB 14; Length 7766;
Best Local Similarity 47.4%; Pred. No. 3.3; Mismatches 111; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 157 GCTAACGCTGCTGCTCTCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 216
DB 4432 GTTACGCTGCTGCTGCTCTGCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 4491
QY 217 AGCGTTATGTATGTAACGCGCGATGTAGCGCCAGGCGGTGGATTAATAGTATTGAAC 276
DB 4492 ATTGCTGCTGCTGCTGCTGCTGCAAGCGATGCCGTAATCTGAACGACCATTAACCG 4551
QY 277 ACTCAGATGTTTCAAGAAATATATCCATCCATCCAGTGAAGCGTAAATACTCCGAT 336
DB 4552 AAGAAAGATGATCAGATTGCTGCTCTATGCTTTGAGGGGATGCTTAAGATGGAAG 4611
QY 337 ATTACTGTGCGCAATAGCGCGTAAATACG 367
DB 4612 TTGCTGTGAAGATGATGATGATGGAAG 4642

RESULT 9

US-09-738-626-482
; Sequence 482, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 482
; LENGTH: 1059
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-482

Query Match 7.2%; Score 32.8; DB 9; Length 1059;
Best Local Similarity 47.2%; Pred. No. 1.8; Mismatches 112; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 184 GATGCCCGTAATCTGAACGACCATTAACGAGCGGTTATGTTAAGCGCGATGTA 243
DB 76 GAGGACGTTTCCACCGCAGACCAACGCAAGCTTCCGCTTAACGATCGATGCA 135
QY 244 GCGCAGGCTGCGATTAATAGTATTGAACCTAGTACAGATGTTTCAAGAAATATGCC 303
DB 136 GCCGCTGAAGAAAGTAAACATCAGCGTCAACCTCTAGAGGAAAGTCAAGAG 195
QY 304 ACCATCGACGATGAAGCACTAAACTCCGATTTATCTGCGCAATAGCGCGTAAAT 363

Db 196 ATCAACAAGGCGTTCATGTAAGCAACCAAGATATTGAGTTGAGTGTACCGCGCTGT 255
Qy 364 AACGCGCGCTGTTAATCAAGCCGATCTGA 395
Db 256 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 287

RESULT 10

US-09-738-626-481/c
Sequence 481, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO
APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 481
LENGTH: 1113
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-481

Query Match 7.2%; Score 32.8; DB 9; Length 1113;
Best Local Similarity 47.2%; Pred. No. 1.9;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATCCCGTAAATTTGAACGACCAATTAACCAAGCGGTTATGTTACGCGCGATGA 243
Db 972 GAGGACGTTTCACCGCAGAGACCAACCACTGCAAGCTCTCCGTTAAGCATCGATGCA 913
Qy 244 GCGCAGGTCGGATTAATGACTATGTAAGTCAAGTCAAGTGTTCAGAAATTAATGCC 303
Db 912 GCGCGTAAAAAGTAAACCATCATCCGCTTCACTTGAGCCTGAGAAAAAGTCGATGAG 853
Qy 304 ACCATGACACGATGGAAGCTTAATACTCGATATTACTGTGCGCAATACGCGGTAAT 363
Db 852 ATCAACAAGGCTTCATGTAAGCAACCAAGATATTAGGTTGAGGTATACCGCGCTGT 793
Qy 364 AACGCGCGCTGTTAATCAAGCCGATCTGA 395
Db 792 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 761

RESULT 11
US-09-738-626-1/c
Sequence 1, Application US/09738626
Publication No. US20020197605A1
GENERAL INFORMATION:
APPLICANT: NAKAGAWA, SATOSHI
APPLICANT: MIZOGUCHI, HIROSHI
APPLICANT: ANDO, SEIKO
APPLICANT: HAYASHI, MIKIRO
APPLICANT: OCHIAI, KEIKO
APPLICANT: YOKOI, HARUHIKO

APPLICANT: TATEISHI, NAOKO
APPLICANT: SENOH, AKIHIRO
APPLICANT: IKEDA, MASATO
APPLICANT: OZAKI, AKIO
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-125
CURRENT APPLICATION NUMBER: US/09/738,626
PRIOR FILING DATE: 2000-12-18
PRIOR APPLICATION NUMBER: JP 99/377484
PRIOR FILING DATE: 1999-12-16
PRIOR APPLICATION NUMBER: JP 00/159162
PRIOR FILING DATE: 2000-04-07
PRIOR APPLICATION NUMBER: JP 00/280988
PRIOR FILING DATE: 2000-08-03
NUMBER OF SEQ ID NOS: 7059
SOFTWARE: PatentIn ver. 3.0
SEQ ID NO 1
LENGTH: 3309400
TYPE: DNA
ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 7.2%; Score 32.8; DB 9; Length 3309400;
Best Local Similarity 47.2%; Pred. No. 1.3e+02;
Matches 100; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

Qy 184 GATCCCGTAAATTTGAACGACCAATTAACCAAGCGGTTATGTTACGCGCGATGA 243
Db 451820 GAGGACGTTTCACCGCAGAGACCAACCACTGCAAGCTCTCCGTTAAGCATCGATGCA 451761
Qy 244 GCGCAGGTCGGATTAATGACTATGTAAGTCAAGTGTTCAGAAATTAATGCC 303
Db 451760 GCGCGTAAAAAGTAAACCATCATCCGCTTCACTTGAGCCTGAGAAAAAGTCGATGAG 451701
Qy 304 ACCATGACACGATGGAAGCTTAATACTCGATATTACTGTGCGCAATACGCGGTAAT 363
Db 451700 ATCAACAAGGCTTCATGTAAGCAACCAAGATATTAGGTTGAGGTATACCGCGCTGT 451641
Qy 364 AACGCGCGCTGTTAATCAAGCCGATCTGA 395
Db 451640 ACTGCGCATCTGACTGCTCGCATTTGAAGCTGA 451609

RESULT 12
US-10-333-631-1
Sequence 1, Application US/10333631
Publication No. US20030190651A1
GENERAL INFORMATION:
APPLICANT: Bayer AG
TITLE OF INVENTION: REGULATION OF HUMAN PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE C
FILE REFERENCE: L10121 Foreign Countries
CURRENT APPLICATION NUMBER: US/10/333,631
PRIOR FILING DATE: 2003-01-23
PRIOR APPLICATION NUMBER: US 60/221,917
PRIOR FILING DATE: 2000-07-31
PRIOR APPLICATION NUMBER: US 60/280,112
PRIOR FILING DATE: 2001-04-02
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 568
TYPE: DNA
ORGANISM: Homo sapiens
FEATURES:
NAME/KEY: misc feature
LOCATION: (90)-(90)
OTHER INFORMATION: n=a, c, g or t
US-10-333-631-1

Query Match 7.1%; Score 32.2; DB 14; Length 568;
Best Local Similarity 52.6%; Pred. No. 2.1;
Matches 70; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 224 ATGTTAACGGCGCCGATGTAGGCCAGGCTCGGATAATAGTACTATTGAACTGACTCAGA 283

Db	4229	ATGATAGTAAAGAGACTTCGCGGACATTGCTCCGACCAAGATGATGACCGCATCACA	4170
Oy	284	ATGCTTTCAGAAATATATCCACCATCGACAGTGAACGCTAAAACTCCGATATTACTG	343
Db	4169	ACGCTGCCACATGCTCGGCTTCGACCAAGGCGCATCGACAGTACACCGATGCTCG	4110
Oy	344	TCGGCCATAGGCGGTAATAACGCGCGG	372
Db	4109	TCGGTCAAGACGATGTCAGCGGCTCTCG	4081

Search completed: March 17, 2004, 08:16:08
 Job time : 411.736 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds

(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgcgtacacagttatcaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : EST:

1:	em_eacba:*
2:	em_eacbm:*
3:	em_eactin:*
4:	em_eactmu:*
5:	em_eactov:*
6:	em_eactpl:*
7:	em_eactro:*
8:	em_hrc:*
9:	gb_eact1:*
10:	gb_eact2:*
11:	gb_hrc:*
12:	gb_eact3:*
13:	gb_eact4:*
14:	gb_eact5:*
15:	em_eactfun:*
16:	em_eactom:*
17:	em_gss_hum:*
18:	em_gss_inv:*
19:	em_gss_pln:*
20:	em_gss_vrt:*
21:	em_gss_fun:*
22:	em_gss_mam:*
23:	em_gss_mus:*
24:	em_gss_pro:*
25:	em_gss_rtd:*
26:	em_gss_rhg:*
27:	em_gss_vrt:*
28:	gb_gss1:*
29:	gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
C 1	223	48.9	680 12	B618688 B618688
C 2	37.4	8.2	484 9	AV430994 AV430994
C 3	37.2	8.2	772 29	CG862779 CG862779
C 4	35.8	7.9	484 14	CF089570 CF089570

5	35.2	7.7	813	28	B2204853
6	34.6	7.6	702	12	BQ441624
7	34.6	7.6	937	12	BQ444255
8	34.2	7.5	296	9	AV058630
9	34.2	7.5	523	12	BJ334624
C 10	34	7.4	811	28	B2469000
C 11	33.8	7.4	480	14	CF372418
C 12	33.8	7.4	481	14	CF414539
13	33.8	7.4	509	14	CF605114
14	33.6	7.4	594	12	BJ335653
15	33.6	7.4	653	12	BJ336964
C 16	33.6	7.4	712	14	CF437982
C 17	33.4	7.3	679	29	CC961526
C 18	33.4	7.3	695	28	BH653951
C 19	33.4	7.3	700	28	BH651371
C 20	33.4	7.3	709	28	BH690914
C 21	33.4	7.3	785	28	BH578793
C 22	33.4	7.3	815	28	BH461716
C 23	33.4	7.3	827	28	BH438448
C 24	33.4	7.3	830	28	BH715216
C 25	33.4	7.3	835	28	BH575194
C 26	33.4	7.3	853	28	BH578402
C 27	33.4	7.3	1026	28	B2458784
C 28	33.2	7.3	589	28	B2300668
29	33.2	7.3	625	10	BF501499
30	33.2	7.3	788	9	AU139321
31	33	7.2	456	12	BJ331339
32	33	7.2	524	12	BJ367372
33	33	7.2	547	12	BJ333958
34	33	7.2	553	12	BJ336903
35	33	7.2	559	12	BJ334047
36	33	7.2	574	12	BJ330328
37	33	7.2	607	12	BJ365766
38	33	7.2	625	12	BJ396922
39	33	7.2	629	12	BJ353378
40	33	7.2	635	12	BJ397391
41	33	7.2	645	12	BJ392753
42	33	7.2	670	12	BJ336637
C 43	32.8	7.2	518	10	BF755518
C 44	32.8	7.2	572	12	B1626794
45	32.8	7.2	574	10	BF854273

ALIGNMENTS

RESULT 1
LOCUS B618688/c 680 bp mRNA linear EST 01-OCT-2003
DEFINITION B618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XL186b22 5', mRNA sequence.
ACCESSION B618688
VERSION B618688.1 GI:37256713
KEYWORDS EST.

SOURCE
ORGANISM Xenopus laevis (African clawed frog)

REFERENCE
AUTHORS Kohara, Y., Kikuchi, A., Terashima, C., Mochii, M., Ueno, N., Shin-i, T. and
1 (bases 1 to 680)
Kilavaya, A., Terashima, C., Mochii, M., Ueno, N., Shin-i, T. and

TITLE
JOURNAL
COMMENT
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasi Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
The information of this clone is available through the following
URL.

FEATURES
source
http://xenopus.nibb.ac.jp.
location/Qualifiers

1. 680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X1186b22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_id="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 48.9%; Score 223; DB 12; Length 680;
Best Local Similarity 68.0%; Pred. No. 3.1e-58;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGCAGCATTCGCGACATTCGAGTTCCTGGCAGTCTTGGCT 60
DB 593 ATGAACCTTTAAAGTGCAGCATTCGCGACATTCGAGTTCCTGGCAGTCTTGGCT 60
QY 61 GCGCTATGATCAGCTGGTACCGGCTGTTGATCCATGAATGACATGATCGGC 120
DB 533 GCGCTATGATCAGCTGGTACCGGCTGTTGATCCATGAATGACATGATCGGC 120
QY 121 CCGGACTAAGCTTATGATTCAGTACGCTTCCGCTAACGCTGCGCTTATATAGCGGC 474
DB 473 CCGGACTAAGCTTATGATTCAGTACGCTTCCGCTAACGCTGCGCTTATATAGCGGC 474
QY 181 AGCGATGCCGCTAAATCTGAACGACATTAACCGAGCGCTTATGTAACCGCGCAT 240
DB 413 ACTGATGCCGCTAAATCTGAACGACATTAACCGAGCGCTTATGTAACCGCGCAT 240
QY 241 GTAGGCCAGCGCTGCGGATTAATGATCTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 353 GTAGGCCAGCGCTGCGGATTAATGATCTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGACGTAAGTGAACGCTAAATCTCGAATTAATGTAATGTAATGTAAT 294
DB 293 GCTACTCTTATGATGAGTGAACGCTAAATCTCGAATTAATGTAATGTAATGTAAT 234
QY 361 AATACGCGCGCTGCTTATGATGAGTGAACGCTAAATCTCGAATTAATGTAATGTAAT 420
DB 233 GCGAAGCGCTGCTTATGATGAGTGAACGCTAAATCTCGAATTAATGTAATGTAAT 174
QY 421 GCTTTGGCAACACCGCGCTAATGTAATGTAATGTAATGTAATGTAATGTAAT 456
DB 173 GCGTTGTATACACGCGCTAATGTAATGTAATGTAATGTAATGTAATGTAAT 138

RESULT 2
AV430994/c 484 bp mRNA linear EST 23-AUG-2000
LOCUS AV430994 Porphyra yezoensis TV-1 Porphyra yezoensis cDNA clone
DEFINITION AV430994 r 5', mRNA sequence.
ACCESSION AV430994
VERSION AV430994.1 GI:8586219
KEYWORDS EST.
SOURCE Porphyra yezoensis
ORGANISM Porphyra yezoensis
Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
Porphyra.
1 (bases 1 to 484)
Nakado, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and
Tabata, S.
Generation of 10,154 expressed sequence tags from a leafy
gametophyte of a marine red alga, Porphyra yezoensis
DNA Res 7, 223-227 (2000)
20363100
PUBMED 10907854

COMMENT
Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/
location/Qualifiers

1. 484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TV-1"
/db_xref="taxon:2788"
/clone="p1027e10_r"
/clone_id="Porphyra yezoensis TV-1"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 2.4;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 CGCTAAACTCCGATATTAATGTCGCGCCATATGCGGCTATATACCGCGCTGTTAA 380
DB 441 CCCAAACACACAGGTCCCGCGCTCAACACCGCTCAATACGCTGCTGGCCAG 382
QY 381 TCAGACGATCTGATTCAGGCTAATGTCGCTGATGCTGTTGGCAACAGCCA 439
DB 381 GCCGAAAGCTGATTCGAAAGTACTGTGCGAGGTCGCGGAAAGGCGCGCGCA 323

RESULT 3
CG862779/c 772 bp DNA linear GSS 19-NOV-2003
LOCUS ZMBB0273B12f ZMBB02 (EcoRI) Zea mays subsp. mays genomic clone
DEFINITION ZMBB0273B12 5', genomic survey sequence.
ACCESSION CG862779
VERSION CG862779.1 GI:38436806
KEYWORDS GSS.
SOURCE Zea mays subsp. mays (maize)
ORGANISM Zea mays subsp. mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
1 (bases 1 to 772)
Zohovetz, V., Fuks, G., Yu, Y., Wing, R. and Messing, J.
Bharti, A.K., Young, S., Kavchok, S., Keizer, G., Bronzino, A.C.,
Sequencing of the maize genome at PGR (2003)
Unpublished (2003)
Contact: Bharti, A.K.
Dr. Joachim Messing's lab
The Plant Genome Initiative at Rutgers, Waksman Institute, Rutgers
University
190 Frelinghuysen Road, Piscataway, NJ 08854, USA
Tel: 732 445 3801
Fax: 732 445 5735
Email: bharti@waksman.rutgers.edu
Seq primer: 17
Class: BAC ends
High quality sequence start: 450.

FEATURES
source
location/Qualifiers

1. 772
/organism="Zea mays subsp. mays"
/mol_type="genomic DNA"
/cultiivar="B73"
/sub_specles="mays"
/db_xref="taxon:4578"
/clone="ZMBB0273B12"
/lab_host="E. coli DH10B"
/clone_id="ZMBB02 (EcoRI)"
/note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI"

ORIGIN

Query Match 8.2%; Score 37.2; DB 29; Length 772;
Best Local Similarity 51.9%; Pred. No. 3.6;
Matches 84; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 45 TGGCAGTCTCTGCTGGCGCTGTATGATCAGCTGGTACCCTGTGTTACCATGAAT 104
DB 202 TGTTCATCAGCTCGCTCCCGCCCGGCGGCGCTTCCGCTCCCACTTGTCTAACAGAGTGGC 143
QY 105 GGCACATCATCCGGCGCCGAGCTCAACGTTGAGCATTAAGTACGGCTTCGCTAACGC 164
DB 142 GGCAGGTGTTCCGCGCAAGACACCGCATGAGCATCTGTGGTATGTTGGCAGCTCGG 83
QY 165 TGGCGTCTCTGCAAGAGCATGCCCGTAAATCTGAACGAC 206
DB 82 TGAGCTGCATTGGAACATTTCAGCAATGTATGTCAGAGAGAC 41

RESULT 4
CF089570/c 484 bp mRNA linear EST 22-JUL-2003
DEFINITION QHM23E04.yg.ab1.QH_M sunflower H. argophyllus Helianthus argophyllus
ACCESSION CF089570
VERSION CF089570.1 GI:33128637
KEYWORDS EST.
SOURCE Helianthus argophyllus
ORGANISM Helianthus argophyllus
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.
1 (bases 1 to 484)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, U., Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lecture and Sunflower ESTs from the Composite Genome Project
http://composita.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore Lab University of California at Davis (UCD) Amundson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659 Email: akozik@ucdavis.org [michelmore@vegmail.ucdavis.edu] belongs to contig QH_Ca_Contig2049, see http://cgpdb.ucdavis.edu/ for details.
Plate: QHM23 row: E column: 04.

FEATURES
source
1. 484
Location/Qualifiers
/organism="Helianthus argophyllus"
/mol_type="mRNA"
/db_xref="taxon:73275"
/clone="QHM23E04"
/lab_host="E.coli"
/clone_lib="QH_M sunflower H. argophyllus"
/note="Vector: pBRCDNA5f1ab; The library was constructed from three different sources (seedling, root and leaf) of RNA from a single genotype. cDNAs were pooled and directionally cloned into a custom medium-copy vector. Details of library construction can be obtained at http://cgpdb.ucdavis.edu/"

ORIGIN
Query Match 7.9%; Score 35.8; DB 14; Length 484;
Best Local Similarity 46.0%; Pred. No. 7.4;
Matches 121; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 58 GCTGGCGCTATGATCAAGCTGTACCGCTGTTGTTACCATGAATGCAATGCAATCC 117
DB 323 GCTGGGTTCAATGATCCCTGATACCGGCTCTGTAAGATGAATGCAATGCAATCC 264
QY 118 GGGCCGCACTAACGTTGAGCATTTATCAAGAGGTTCCGTTAACGCTGCGCTTGTCTG 177
DB 263 GATCTCTAGTCGCAATCCGGGCTGCTCTCAAGGCTTGAATCTGTAGTACGCGCTGAA 204

QY 178 CAAGCGATGCGCCGTAATCTGAACGACCACTTACCAGCGGTTATGTATGAGCGGCC 237
DB 203 GTGACAAACATCATTTAGCAAAAGTTTCAATGATCAAGAGCTTGTACTCTGTC 144
QY 238 GATGTAGGCGCAGGCTGCGATTAATGACTATTGAACTGACTCAGATGTTTCAAAAT 297
DB 143 CCAGAGGCGGTGTTGTGCTTCAATTTGTTATGATCAATTAATCGTACCGCAAT 84
QY 298 AATGCCACCATCAGACCACTGAA 320
DB 83 GAAGCCAAAGATGAGACCTGTGAA 61

RESULT 5
BZ204853
LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21_TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
ACCESSION BZ204853
VERSION BZ204853.1 GI:23862905
KEYWORDS GSS.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 813)
Zhao, S., Shetty, J., Shatsman, S., Tesgaye, G., Geer, K., Shvartsbeyn, A., Gebregorgis, B., Overton, L., Russell, D., Chen, D., Riggs, F., de Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 MboI segment
Unpublished (1999)
Other_GSSs: CH230-390D21_TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering_information.htm). BAC end
page: http://www.tigr.org/cdb/bac_ends/rat/bac_end_intro.html
Seq primer: T7
Plate: 390 row: D column: 21
Class: BAC ends.

FEATURES
source
1. 813
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/strain="BN/SNHed/MCW"
/db_xref="taxon:10116"
/clone="CH230-390D21"
/sex="Female"
/cell_type="Brain"
/clone_lib="CHORI-230 Segment 2"
/note="Vector: pTARAC1.3; Site 1: MboI; Site 2: MboI; CHORI-230 Rat (BN/SNHed/MCW) BAC library produced by Pieter de Jong"

ORIGIN
Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 16;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 178 CAAGCATGCGCCGTAATCTGAACGACCACTTACCAGCGGTTATGTATGATGAGCGGCC 237
DB 548 CAAGGACCTCACAATCAACGACACACTCAACTAATGAATAAAGAACTTGAAGGAGCA 607
QY 238 GATGTAGGCGCAGGCTGCGATTAATGACTATTGAACTGACTCAGATGTTTCAAAAT 297

Db 608 TCTGGACACATGGGCACTGGAAAAAATCTCTGACAAAACACCAATGCTTATCTCT 667
 QY 298 AATGCCACCATCGACCACTGGAACGCTTAAATCTCCGATTTATCTGCGCCCAATACGGC 667
 Db 668 AAGATCAATATGACCAATGAGTCTCTAATTAATGCAACAACTTCTGAGCAAGACAC 357
 QY 358 GGTAAATACGCC 369
 Db 728 TGTGTTAGGAC 739

RESULT 6
 BG441624
 LOCUS
 DEFINITION BG441624 702 bp mRNA linear EST 15-MAR-2001
 accession GA_Ea0014A14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 version BG441624
 keywords BG441624.1 GI:13351276
 source EST.
 organism Gossypium arboreum
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 702)
 Henry, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Wing, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)
 contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATAGGG
 High quality sequence stop: 696.
 Location/Qualifiers
 1..702
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0014A14f"
 /tissue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

FEATURES
 source

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 702;
 Best Local Similarity 46.5%; Pred. No. 22;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 181 AGCGATGCCCGTAATCTGAACGACCATTAACCAAGCGGTTATGTAACGGCGCGAT 240
 Db 185 AACATGACGGGAATGCGAATGGGATTTGGACACACCAATCAAGAGACACCAAT 244
 QY 241 GTAGCCAGGGTGGGATTAATGTAATTAAGTAACTGACTCAGATGGTTTCAAAATAT 300
 Db 245 GACAATGTAATCAATGAATGTAATCTCAAAAGAAAAACAAGATTCAGTTCAAT 304
 QY 301 GCCACCATGACGACGTGAACGCTTAAATCTCCGATTTATCTGCGCCCAATACGGCGGT 360
 Db 305 GAAATGCTGACAGGACCAACCAACAAATGAATAAACAATTAACATGAAATGCGGT 364
 QY 361 AATAAGCCGCGCTGTTATCAAGCCGCACTTGATTCAGGCTATGTTGCTCAGTT 420

Db 365 CAGACGGGAGAAATGCCACTGAGATTAACACTGACAAACAATGAATAATGTTGACGAAT 424
 QY 421 G 421
 Db 425 G 425

RESULT 7
 BG444255
 LOCUS
 DEFINITION BG444255 937 bp mRNA linear EST 15-MAR-2001
 accession GA_Ea0023M14f Gossypium arboreum 7-10 dpa fiber library Gossypium
 version BG444255
 keywords BG444255.1 GI:13353907
 source EST.
 organism Gossypium arboreum
 Gossypium arboreum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eustosids II; Malvales; Malvaceae; Malvoideae; Gossypium.
 1 (bases 1 to 937)
 Henry, R.A., Fritsch, D., Yu, Y., Main, D., Rambo, T., Simmons, J.,
 Wing, D., Wood, T.C., Leslie, A. and Wilkins, T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)
 contact: Wing RA
 Clemson University Genomics Institute
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: twing@clemson.edu
 Seq primer: TAATACGACTCACTATAGGG
 High quality sequence stop: 804.
 Location/Qualifiers
 1..937
 /organism="Gossypium arboreum"
 /mol_type="mRNA"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_Ea0023M14f"
 /tissue_type="Fibers isolated from bolls harvested 7-10
 dpa"
 /lab_host="E. coli"
 /clone_1lb="Gossypium arboreum 7-10 dpa fiber library"
 /note="Vector: pBR-CMV; Site_1: EcoRI; Site_2: XhoI"

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

FEATURES
 source

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 937;
 Best Local Similarity 46.5%; Pred. No. 26;
 Matches 112; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 181 AGCGATGCCCGTAATCTGAACGACCATTAACCAAGCGGTTATGTAACGGCGCGAT 240
 Db 185 AACATGACGGGAATGCGAATGGGATTTGGACACACCAATCAAGAGACACCAAT 244
 QY 241 GTAGCCAGGGTGGGATTAATGTAATTAAGTAACTGACTCAGATGGTTTCAAAATAT 300
 Db 245 GACAATGTAATCAATGAATGTAATCTCAAAAGAAAAACAAGATTCAGTTCAAT 304
 QY 301 GCCACCATGACGACGTGAACGCTTAAATCTCCGATTTATCTGCGCCCAATACGGCGGT 360
 Db 305 GAAATGCTGACAGGACCAACCAACAAATGAATAAACAATTAACATGAAATGCGGT 364
 QY 361 AATAAGCCGCGCTGTTATCAAGCCGCACTTGATTCAGGCTATGTTGCTCAGTT 420
 Db 421 G 421
 QY 425 G 425

RESULT 8
LOCUS AV058630
DEFINITION AV058630 Mus musculus pancreas C57BL/6J adult Mus musculus cDNA clone 1810054102, mRNA sequence.
ACCESSION AV058630
VERSION AV058630.1 GI:5158377
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE Carinetti, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K., Akahira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T., Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M., Kawai, D., Kikuchi, N., Kojima, Y., Matsuyama, T., Nitsuna, H., Oda, H., Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Sogabe, Y., Sugihara, Y., Suzuki, H., Suzuki, H., Tateno, M., Tomaru, Y., Tomimaga, N., Watanabe, S., Tagame, M., Yamamura, T., Yokota, T., Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.
TITLE RIKEN Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@rtc.riken.go.jp
Thermotabilization and thermostabilization of thermostable enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES
source location/Qualifiers
1..296
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
/clone_lib="Mus musculus pancreas C57BL/6J adult"

ORIGIN
Query Match 7.5%; Score 34.2; DB 9; Length 296;
Best Local Similarity 53.3%; Pred. No. 17;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 251 GTCGGATTAATGCTACTATGACGACACGAATGCTTTCGAATAATGCCACCATCG 310
DB 13 GTTGAGACCCAGTCTTTCTGAGTCACCAAGCTGGGGTAAAGAAAGACATCAACG 72
QY 311 ACCAGTGAACGCTAAACCTCGATATTACTGTGGCAATACGCGGTATTAACGCCG 370
DB 73 TCCAGTGAAGCAACAAAATGAAGAAATGATTATCCCAAAAGTTGAATTAAAGTAG 132
QY 371 CGCTGTTAATCAGA 385
DB 133 AGATGCCAAATATAGA 147

RESULT 9
LOCUS BJ334624
DEFINITION BJ334624 Dictyostellium discoideum cDNA library, AF Dictyostellium

discoideum cDNA clone dda47006 5', mRNA sequence.
ACCESSION BJ334624
VERSION BJ334624.1 GI:19164754
KEYWORDS EST.
SOURCE Dictyostellium discoideum
ORGANISM Dictyostellium discoideum
Eukaryota; Mycetozoa; Dictyostelidia; Dictyostelium.
REFERENCE Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostellium discoideum at the aggregation stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadao Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tsbinfo@nig.ac.jp.

FEATURES
source location/Qualifiers
1..523
/organism="Dictyostellium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="dda47006"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostellium discoideum cDNA library, AF"

ORIGIN
Query Match 7.5%; Score 34.2; DB 12; Length 523;
Best Local Similarity 57.1%; Pred. No. 25;
Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 238 GATGTAGCCACGAGGTGGGATATGACTATTGAAGTGAAGTCAAGATGTTCAAGAAAT 297
DB 398 GATGAAGATGATGATTTGGAGATGATGATATGGAATCAATGGAATGACATTAAT 457
QY 298 AATGCCACCATGACCAAGTGAACGCTAAACCTCCGATTACT 342
DB 458 AATAACAAACACACACAAATATATATAAATAAATAATTAAT 502

RESULT 10
LOCUS B2469000/c
DEFINITION B2469000/c 811 bp DNA linear GSS 13-DEC-2002
B2469000/c BOA058TF BO.1.6.2 KB tot Brassica oleracea genomic clone BOA058, genomic survey sequence.
ACCESSION B2469000
VERSION B2469000.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta, eudicotyledons; core eudicots; rosids; eustrodi II; Brassicales; Brassicaceae; Brassica.
REFERENCE Tom, C.D., Van Aken, S., Uteirback, T., Koo, H. and Fraser, C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other_GSS: BOA058TR
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523
Fax: 301-838-0208
Email: cdtown@tigr.org
DNA is from a doubled haploid provided by Tom Osborn.
Seq primer: TP
Class: sheared ends.
FEATURES
source location/Qualifiers
1..811

```
/clone_lib="BO_1.6_2 KB tot"
/notes="vector: pHSOI; Site_1: BstXI; 1.6-2 kb sheared
total DNA inserted into pHSOI using BstXI linkers"
```

Query Match	7.5%;	Score 34;	DB 28;	Length 811;
Best Local Similarity	48.9%;	Pred. No. 37;		
Matches 91;	Conservative	0.1	Mismatches 65	

	176	610	236	550	296	490	356	430	425
Qy	TCGAAAGCATCCCCGTTAAATCTGTAAGAACGACCATTAACAGACGGTATGTGTAACGGCG								
Db									
	610	TGTTAGGCAAGCATGTGTAATAAATCTGTCTCCCATTCATCTACTGTTTATGCAAAATCT							
Qy	236	CCGATGTAGGCGAGGGTCCGATTAATGTACTTGAATGACTGACGATGGTTTCAGAA							
Db	550	TGTATCTAATTAACCTGCTGATTTTGGACTCTCTCTGTTGTCGTGGAGATAGTTTAAAT							
Qy	296	ATAAGCCACCATGACCAACGAGCGCTTAAATCTCCATATTAATCTCGGCCAATACG							
Db	490	ACTTTGTACACTTATTAATAAGAAAAATCGAAAAATACACTGATTAACACTTATTCAAACCA							
Qy	356	GCGGTA 361							
Db	430	AAAGTA 425							

RESULT 11	CF372418/c	LOCUS	DEFINITION	CF372418	480 bp	mRNA	linear	EST 27-AUG-2003
			CSEC052B08, F10N0012 Cabsau Normalised F10N0012					
			Vitis vitifera CDNA clone CSEC052B08 3', mRNA sequence.					
			CF372418					

VERSION
 CF372418.1
 KEYWORDS
 EST
 SOURCE
 GI:34319664
 ORGANISM
Vitis vinifera
Vitis vinifera
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliopsida; euarchaeplastids, core eudicots; Rosids; Vitaceae; Vitis.
 REFERENCE
 1 (bases 1 to 480)
 AUTHORS
 Iocco, P., Hua, C., Davies, C. and Thomas, M.R.
 TITLE
 Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon
 JOURNAL
 Unpublished (2003)
 COMMENT
 Contact: Mark R. Thomas

FEATURES
 source Location/Qualifiers
 1 480
 CSIRO Plant Industry
 CSIRO
 PO Box 350, Glen Osmond, SA, 5064, Australia
 Tel: 61 8 83038600
 Fax: 61 8 83038601
 Email: Mark.R.Thomas@csiro.au
 Seq Primer: CCCACTCAGCAGTGTGAAACG (M13 Forward)
 POLVA=yes

```

/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSECS052B08"
/sex="Hermaphrodite"
/dev_stage="12 - modified E-L system"
/clone_1b="CabSau Normalised Flower Stage 12 (FL0n0012)"
/note="Organ: Inflorescence including flowers; Vector:
pZ1; Normalised cDNA library from immature inflorescences
at stage 12 of the modified E-L system. Tissue collected
from field grown plants. A description of the modified E-L
system can be found in the paper by B. G. Coombe 'Adoption

```

[illegible]

of a system for identifying grapevine growth stages (1995) Aust J Grape and Wine Res.

RESULT	12
CP414539/c	
LOCUS	
DEFINITION	481 bp mRNA linear EST 02-SEP-2003
ACCESSION	CSEC6071B09 POS00036 CabSau Normalised Berry Postveraison Stage 36
VERSION	CP414539
KEYWORDS	Vitis vinifera cDNA clone CSEC6071B09 3', mRNA sequence.
EST	CP414539.1 GI:34415785
SOURCE	
ORGANISM	Vitis vinifera
	Vitis vinifera

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosids; Vitaceae; Vitis.				
1 (bases 1 to 481)				
Iocco, P., Hua, C., Davies, C. and Thomas, M.R.				
Expressed sequence tags from the grapevine cultivar Cabernet Sauvignon				
Unpublished (2003)				
Contact: Mark R. Thomas				
CSIRO Plant Industry				
CSIRO				
PO Box 350, Glen Osmond, SA, 5064, Australia				
Tel: 61 8 83038600				
Fax: 61 8 83038601				
Email: Mark.R.Thomas@csiro.au				
Seq primer: CCCAGTCACGACGTTGTAACAAG (M13 Forward)				
POLYA=yes.				

```

FEATURES
source
location/Qualifiers
1..481
/organism="Vitis vinifera"
/mol_type="mRNA"
/cultivar="Cabernet Sauvignon"
/db_xref="taxon:29760"
/clone="CSEC5071B09"
/sex="Hermaphrodite"
/dev_stage="36 - modified E-L system"
/clone_lib="Cabsau Normalised Berry Postveraison Stage 36 (posn036)."
/note="Organ: Fruit with seeds removed; Vector: pZU; A cDNA library from postveraison fruit at stage 36 of the modified E-L system. Tissue collected from field grown plants. A description of the modified E-L system can be found in the paper by B. G. Coombe 'Adoption of a system for identifying grapevine growth stages' (1995) Aust. J. Grape and Wine Res. 1: 104-110."

```

Query Match	7.4%	Score 33.8;	DB 14;	Length 481;
Best Local Similarity	49.7%	0;	Match No. 31;	
Matches 86;	Conservative	0;	Mismatches 87;	Indels 0;
Gaps 0;				

Db 223 ATCAATGAGGTGGAACCTGACCTTCCCTCCCGCAACCACTGTGGAGAGGGCGCT 164
 Qy 154 TCCGCTACGCTGCGCTTCTCTGCAAGCGATGCCGTAATTCGAAACACCATTTACC 213
 Db 163 GCCGCTGCGCTGCGGCTCTCTGCTGCTAGCAAGACACAGGCGCTAAAGCCACATTTCC 104
 Qy 214 CAGAGCGGTATGTATGTAAGCGCGCGCATGTAGCGCGGCGGATTAATAGTAC 266
 Db 103 CAGAAAGATCAAGGCTTGAAAGCAGCAGTACCGCGGATCAGATGATTCTCC 51

RESULT 13
 CF605114 509 bp mRNA linear EST 30-SEP-2003
 LOCUS RADIC01_000700 Grape Root PSPORT1 Library Vitis vinifera cDNA 5',
 DEFINITION mRNA sequence.
 ACCESSION CF605114
 VERSION CF605114.1 GI:37185761
 KEYWORDS EST.
 SOURCE Vitis vinifera
 ORGANISM Vitis vinifera

REFERENCE 1 (bases 1 to 509)
 Moser, C., Segala, C., Fontana, P., Salakhudinov, I., Gatto, P., Pindo, M., Zyprian, E., Toepfer, R., Grandi, M. S. and Velasco, R.
 Title Pindolo, a new Vitis vinifera variety from the Piedmont region (Italy).
 Journal Unpublished (2003)
 COMMENT The sequencing work has been funded by the 'Fondazione Cassa di
 Riparmio di Trento e Rovereto'.
 High quality sequence stop: 509.
 Location/Qualifiers
 1..509
 /organism="Vitis vinifera"
 /mol_type="mRNA"
 /cultiVar="Pinot noir"
 /db_xref="taxon:29760"
 /sex="Hermaphrodite"
 /dev_stage="young root (first year)"
 /lab_host="DH10B"
 /clone_lib="Grape Root PSPORT1 Library"
 /note="Organ: root; Vector: PSPORT1; Site_1: NotI; Site_2: SalI"

FEATURES

source

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 509;
 Best Local Similarity 49.7%; Pred. No. 33;
 Matches 86; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 Qy 94 ACCCATGAATGGAATGCAATGATCGCGCGGAGCTCAAGCTTGAGATTATTCAGTACGAT 153
 Db 295 ATCAATGAGGTGGAACCTGACCTTCCCTCCCGCAACCACTGTGGAGAGGGCGCT 354
 Qy 154 TCCGCTACGCTGCGCTTCTCTGCAAGCGATGCCGTAATTCGAAACACCATTTACC 213
 Db 355 GCCGCTGCGCTGCGGCTCTCTGCTGCTAGCAAGACACAGGCGCTAAAGCCACATTTCC 414
 Qy 214 CAGAGCGGTATGTATGTAAGCGCGCGCATGTAGCGCGGCGGATTAATAGTAC 266
 Db 415 CAGAAAGATCAAGGCTTGAAAGCAGCAGTACCGCGGATCAGATGATTCTCC 467

RESULT 14
 BU335653

LOCUS BU335653 594 bp mRNA linear EST 05-MAR-2002
 DEFINITION BU335653 Dictyostelium discoideum cDNA library, AF Dictyostelium
 accession BU335653
 VERSION BU335653.1 GI:19165783
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
 REFERENCE 1 (bases 1 to 594)
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadashi Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..594
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AY4"
 /db_xref="taxon:44689"
 /clone="ddas1h06"
 /sex="mat A"
 /dev_stage="Aggregation stage"
 /clone_lib="Dictyostelium discoideum cDNA library, AF"

ORIGIN

Query Match 7.4%; Score 33.6; DB 12; Length 594;
 Best Local Similarity 57.1%; Pred. No. 41;
 Matches 60; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Qy 238 GATGAGGCGGAGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 297
 Db 223 GATGAGGCGGAGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGAT 282
 Qy 298 AATGCCACCATGCAACGATGGAACGCTTAATACTCCGATATTACT 342
 Db 293 AATGACACACACACACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 327

RESULT 15
 BU336964

LOCUS BU336964 653 bp mRNA linear EST 05-MAR-2002
 DEFINITION BU336964 Dictyostelium discoideum cDNA library, AF Dictyostelium
 accession BU336964
 VERSION BU336964.1 GI:19167094
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.

REFERENCE 1 (bases 1 to 653)
 AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
 TITLE Full length cDNA of Dictyostelium discoideum at the aggregation
 stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadashi Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1..653
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"

FEATURES
 source

```

/srain="AX4"
/db_xref="taxon:44689"
/clone="ddas5e20"
/sex="mat A"
/dev_stage="Aggregation stage"
/clone_lib="Dictyostelium discoideum cDNA library, Af"

```

ORIGIN

```

Query Match      7.4%; Score 33.6; DB 12; Length 653;
Best Local Similarity 53.9%; Pred.No. 43;
Matches 69; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 238 GATGTAGGCCAGGCGGATATAGTACTATTGAACTGACTCAGAAATGTTGAGAAAT 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 GATGAAGATGATGAATTGAGAAATGATATATTGGAATCAATAGAAATGAGACATATAAT 246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 298 AATGCCACCATGACCACTGAGAACGCTTAAACTCCGATATTACTGTGGCCATATACGC 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 AATTAACAACACACACATATAATATAAATAATATTATATAAGATGATGAAGAA 306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 358 GGTAAATAA 365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 AATAAAAA 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: March 16, 2004, 04:29:07
 Job time : 2232.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds

(without alignment) 10077.856 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcg.....ccacgctcaccagatctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl.*

1: gb_ba.*

2: gb_bcg.*

3: gb_in.*

4: gb_om.*

5: gb_ov.*

6: gb_pat.*

7: gb_ph.*

8: gb_pl.*

9: gb_pr.*

10: gb_ro.*

11: gb_ats.*

12: gb_gy.*

13: gb_un.*

14: gb_vl.*

15: em_ba.*

16: em_fun.*

17: em_hum.*

18: em_in.*

19: em_mu.*

20: em_om.*

21: em_or.*

22: em_ov.*

23: em_pat.*

24: em_ph.*

25: em_pl.*

26: em_ro.*

27: em_ats.*

28: em_un.*

29: em_vl.*

30: em_hcg_hum.*

31: em_hcg_inv.*

32: em_hcg_other.*

33: em_hcg_mus.*

34: em_hcg_pln.*

35: em_hcg_rtd.*

36: em_hcg_mam.*

37: em_hcg_vrt.*

38: em_ay.*

39: em_hgo_hum.*

40: em_hgo_mus.*

41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	406.4	89.1	456	6	I44909	I44909 Sequence 58
2	406.4	89.1	2067	1	SEU43280	U43280 Salmonella
3	400	87.7	5103	1	STPA2301	AJ002301 Salmonella
4	400	87.7	22411	1	AE008749	AE008749 Salmonella
5	398.4	87.4	254050	1	AL627269	AL627269 Salmonella
6	398.4	87.4	301983	1	AE016840	AE016840 Salmonella
7	392.6	86.1	1048	1	STAGEFA	AJ000541 Salmonella
8	282.2	61.9	361	6	I44908	I44908 Sequence 56
9	280	61.4	2889	1	CSP515700	AJ515700 Citrobacter
10	245.2	53.8	2920	1	CFR515701	AJ515701 Citrobacter
11	232	50.9	4680	1	BCCSGABDG	X90754 E. coli c99g
12	232	50.9	10346	1	AE000205	AE000205 Escherichia
13	232	50.9	15047	1	D90741	D90741 Escherichia
14	230.4	50.5	456	6	AX814811	AX814811 Sequence
15	230.4	50.5	648	1	ECOCGGA	L04979 Escherichia
16	220.6	48.4	306358	1	AE016759	AE016759 Escherichia
17	219	48.0	1711	1	AF275733	AF275733 Escherichia
18	219	48.0	10190	1	AE005315	AE005315 Escherichia
19	219	48.0	327773	1	AF002554	AF002554 Escherichia
20	192	42.1	2883	1	ESAS15702	AJ515702 Enterobacter
21	179.4	39.3	230	1	SEU53207	U53207 Salmonella
22	163.2	35.8	10370	1	AE015131	AE015131 Shigella
23	163.2	35.8	292504	1	AE016981	AE016981 Shigella
24	140.2	30.7	437	1	AF237726	AF237726 Shigella
25	93.2	20.4	19201	1	D90742	D90742 Escherichia
26	64.6	14.2	1212	1	EC0131756	AJ3131756 Escherichia
27	48.2	10.6	78	6	AX814809	AX814809 Sequence
28	39.4	8.6	10999	1	AE002536	AE002536 Neisseria
29	39.4	8.6	349980	6	AX044034	AX044034 Sequence
30	37.8	8.3	349061	1	NMA222491	AI162753 Neisseria
31	37	8.1	9955	1	AE002504	AE002504 Neisseria
32	37	8.1	349980	6	AX044033	AX044033 Sequence
33	36.8	8.1	126062	2	AP006135	AP006135 Lotus cor
34	36.6	8.0	110000	2	AC142602	AC142602 Macaca mu
35	36.6	8.0	305730	2	AC114347	AC114347 Rattus no
36	36.6	8.0	311823	1	AC095362	AC095362 Rattus no
37	36.4	8.0	10798	1	AE004604	AE004604 Pseudomon
38	36.4	8.0	110000	2	AC110929_4	AC110929 (5 of
39	36.4	8.0	181996	2	AC108614	AC108614 Rattus no
40	36.4	8.0	212174	2	AC105896	AC105896 Rattus no
41	36.4	8.0	305724	2	AC128763	AC128763 Rattus no
42	36.2	7.9	1797	3	LEIMSP52A	L19563 Leishmania
43	36.2	7.9	1860	3	LIGP63GEN	Z83677 L. infantum
44	36.2	7.9	2104	3	LD0495002	AJ495002 Leishmani
45	36.2	7.9	2106	3	LD0495006	AJ495006 Leishmani

ALIGNMENTS

RESULT 1

LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997

DEFINITION Sequence 58 from patent US 5635617.

ACCESSION I44909

VERSION I44909.1 GI:2469622

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 456)

AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.

TITLE Methods and compositions comprising the agfa gene for detection of Salmonella

JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES
 source 1 Location/Qualifiers
 1.456 /organism="unknown"
 /mol_type="unassigned DNA"

ORIGIN
 Query Match 89.1%; Score 406.4; DB 6; Length 456;
 Best Local Similarity 93.2%; Pred. No. 1.9e-103;
 Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGGAGCATTTGAGCATGCTAGTTCTGGAGCTGCTGGCT 60
 Db 1 ATGAACTTTTAAAGTGGAGCATTTGAGCATGCTAGTTCTGGAGCTGCTGGCT 60
 Qy 61 GGGCTGTTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 60
 Db 61 GGGCTGTTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
 Qy 121 CCGGACTATATACAGCTGGTTACCGGTGTTTACCCATGAAATGGGACATGCTGCA 180
 Db 121 CCGGACTATATACAGCTGGTTACCGGTGTTTACCCATGAAATGGGACATGCTGCA 180
 Qy 181 AGCGATGCCCGTAAATCTGAAACGACATTACCGAGCCGTTATGTTAACGGGCGCAT 240
 Db 181 AGCGATGCCCGTAAATCTGAAACGACATTACCGAGCCGTTATGTTAACGGGCGCAT 240
 Qy 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 Db 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
 Qy 301 GCCACCATGACCACTGAGCACTGTAATTAATCTGTCGGCCATACGGCGGT 360
 Db 301 GCCACCATGACCACTGAGCACTGTAATTAATCTGTCGGCCATACGGCGGT 360
 Qy 361 AATAAGCGCGGCTGTTAATCAACCGCATCTGATTCAGCGTATGCGGTGAGGT 420
 Db 361 AATAAGCGCGGCTGTTAATCAACCGCATCTGATTCAGCGTATGCGGTGAGGT 420
 Qy 421 GGTTTGGCAACACCGGCGGCTAACCACTATTA 456
 Db 421 GGTTTGGCAACACCGGCGGCTAACCACTATTA 456

RESULT 2
 SEU43280. 2067 bp DNA linear BCT 14-FEB-1996
 LOCUS
 DEFINITION Salmonella enteritidis agfBAC operon: fimbria-like protein
 AgfC (agfC) genes, complete cds.
 ACCESSION U43280.1 GI:1184712
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Salmonella enteritidis
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Salmonella.
 1 (bases 1 to 2067)
 Dorian, J.L., Collinson, S.K., Burian, J., Santos, G., Todd, E.C.,
 Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I., and Kay, W.W.
 DNA-based diagnostic tests for Salmonella species targeting agfA,
 the structural gene for thin, aggregative fimbriae
 J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
 JOURNAL MEDLINE 94013373
 PUBMED 8104955
 2 (bases 1 to 2067)
 Collinson, S.K., Clouthier, S.C., Dorian, J.L., Baner, P.A. and
 Kay, W.W.
 Salmonella enteritidis agfBAC operon encoding thin, aggregative
 fimbriae
 J. Bacteriol. 178 (3), 662-667 (1996)
 JOURNAL MEDLINE 9614512
 PUBMED 8550497
 3 (bases 1 to 2067)

AUTHORS Collinson, S.K., Dorian, J.L., Baner, P.A. and Kay, W.W.
 TITLE Direct Submission
 JOURNAL Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
 Microbiology, University of Victoria, P.O. Box 3055 Petch Bldg.,
 Victoria, BC V8W 3P6, Canada
 Location/Qualifiers
 1. 2067
 /organism="Salmonella enteritidis"
 /mol_type="genomic DNA"
 /strain="27655-3b"
 /db_xref="taxon:592"
 /map="between puta and pyrc"
 571. 576
 598. 603
 696. 1151
 /gene="agfB"
 696. 1151
 /gene="agfB"
 /note="fimbria-like protein precursor"
 /codon_start=1
 /product="AgfB precursor"
 /protein_id="AAC43598.1"
 /db_xref="GI:1184713"
 /translation="MKKLLFMMLTILGAPGATATATNYDLARSEYFPAVELSKSPN
 STSQSNVGNRAITIQSGNKNITVGTQTAIVVQKSHMIRVQR"
 696. 752
 /gene="agfB"
 753. 1148
 /gene="agfB"
 /product="AgfB"
 /note="putative signal sequence of 21 amino acids;
 fimbria-like protein"
 1193. 1648
 /gene="agfA"
 1193. 1648
 /gene="agfA"
 /note="thin aggregative fimbria subunit precursor; major
 fimbrial subunit of thin aggregative fimbriae precursor"
 /codon_start=1
 /product="AgfA fimbria precursor"
 /protein_id="AAC43599.1"
 /db_xref="GI:1184714"
 /translation="MDLKVAAFAIVVSGALAGVPOKGGGNNHGGNSGPDST
 LSIYQGSANALALQSDARKSETTITGSGYGNALDVGQADNSTIELTQNGFRNAT
 IDQNAKNSDITVGVGYNALVNGTASDSVVRQVGFNNATANQY"
 1292. 1354
 /gene="agfA"
 1352. 1645
 /gene="agfA"
 /note="AgfA fimbria"
 /product="AgfA fimbria"
 /note="thin aggregative fimbria subunit; major fimbrial
 subunit of thin aggregative fimbriae"
 1667. 1696
 1710. 2036
 /gene="agfC"
 1710. 2036
 /gene="agfC"
 /function="unknown"
 /codon_start=1
 /transl_table=11
 /product="AgfC"
 /protein_id="AAC43600.1"
 /db_xref="GI:1184715"
 /translation="MTLLILALSNQITFTTQGDIVTVIPQVTLNPPCVQVQIL
 SYRQVGGSHQKQTLSPANQPLELRSLSVNISSEDSVXLIIVTVSDGSLHLSQO
 MPBQAC"

ORIGIN
 Query Match 89.1%; Score 406.4; DB 1; Length 2067;
 Best Local Similarity 93.2%; Pred. No. 2.1e-103;

Matches	425; Conservative	0; Mismatches	31; Indels	0; Gaps	0;
QY	1	ATGAACTTTTAAAGTGGCAGCATTCGCAATGTTGTTCTGGCAGTCTGGCT	60		
DB	1193	ATGAACTTTTAAAGTGGCAGCATTCGCAATGTTGTTCTGGCAGTCTGGCT	1252		
QY	61	GGCGTCTTCCCAATGGGCGGCGGTTATCATTAACGGGCGGCAATAGTTCCGGC	120		
DB	1253	GGCGTCTTCCCAATGGGCGGCGGTTATCATTAACGGGCGGCAATAGTTCCGGC	1312		
QY	121	CCGACATTAATCAAGCTGGTTACCCGTTGTTTACCATTAATGACATGCATGCA	180		
DB	1313	CCGACATTAATCAAGCTGGTTACCCGTTGTTTACCATTAATGACATGCATGCA	1372		
QY	181	AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCCGAT	240		
DB	1373	AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCCGAT	1432		
QY	241	GTAGGCGAGGGTGGCGATTAATAGTACTATTGAATGAACTGAGATGGTTTCAGAATAT	300		
DB	1433	GTAGGCGAGGGTGGCGATTAATAGTACTATTGAATGAACTGAGATGGTTTCAGAATAT	1492		
QY	301	GGCACCATTGACAGTGAACGCTAAATCTCCGATATTCTGCGCCCAATACGGCGGT	360		
DB	1493	GGCACCATTGACAGTGAACGCTAAATCTCCGATATTCTGCGCCCAATACGGCGGT	1552		
QY	361	ATAACGCGCGCGTGTATTAATGACACCGCATCTGATCCAGCGTAATGGTGGTCAAGTT	420		
DB	1553	ATAACGCGCGCGTGTATTAATGACACCGCATCTGATCCAGCGTAATGGTGGTCAAGTT	1612		
QY	421	GGTTTGGCAACAGCCGCGGCTAACAGTATTAA	456		
DB	1613	GGTTTGGCAACAGCCGCGGCTAACAGTATTAA	1648		
RESULT 3	STA2301	5103 bp	DNA	linear	BCT 15-NOV-2000
LOCUS	Salmonella typhimurium csgC, csgF, csgE, csgD, csgB, csgA, and csgG genes.				
ACCESSION	AJ002301.1	GI:2739232			
VERSION					
KEYWORDS	csgC gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;				
SOURCE	Salmonella typhimurium				
ORGANISM	Salmonella typhimurium				
REFERENCE	1 Romling, U., Bian, Z., Hammar, M., Sierralta, W.D. and Notmark, S.				
AUTHORS	Cultil fibers are highly conserved between Salmonella typhimurium				
TITLE	and Escherichia coli with respect to operon structure and				
JOURNAL	regulation				
REFERENCE	J. Bacteriol. 180 (3), 722-731 (1998)				
MEDLINE	98117058				
PUBMED	9457880				
REFERENCE	2 (bases 1 to 5103)				
AUTHORS	Romling, U.				
TITLE	Direct Submission				
JOURNAL	Submitted (29-OCT-1997) Romling U., Department of Bacteriology,				
REFERENCE	Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN				
FEATURES	Location/Qualifiers				
source	1. .5103				
	/organism="Salmonella typhimurium"				
	/mol_type="genomic DNA"				
	/strain="SR-11"				
	/db_xref="taxon:602"				
	<1..322				
	/note="ORF179a"				
	/codon_start=2				
	/transl_table=11				
	/product="hypothetical protein"				
	/protein_id="CAA05311.1"				
CDS					
	gene				
	-10_signal				
	-35_signal				
	/db_xref="GI:2739233"				
	/db_xref="GOA:O54290"				
	/db_xref="SWISS-PROT:O54290"				
	/translation="CTLSGMVALVVIHGSALAPHLIVSVYLTGIVAFIMCIGAKO				
	LILSFVPFTFGACATFAGQGDWRVLVLSLALGLIFGYAMKNSGLMLASRRQHSANT				
	AVTK"				
	complement(361..1194)				
	/gene="csgC"				
	complement(361..1194)				
	/gene="csgC"				
	/function="lipoprotein"				
	/codon_start=1				
	/evidence=experimental				
	/transl_table=11				
	/product="Csgg protein"				
	/protein_id="CAA05312.1"				
	/db_xref="GI:2739234"				
	/db_xref="GOA:O54291"				
	/db_xref="SWISS-PROT:O54291"				
	/translation="MPRLITVAVVLLSGCLTAPPKQAKPTLMPRAOSYKDLTLPA				
	FTGKIFSVVNIQDETGFQKPYSPASNSTAVPQSHVLAALVTLAKDSRNFILDERGLQ				
	NLNERKTIIRAQENGTVAMNNRIPLQSLTANIMVEGSIIGYENVSIGGVAFPG				
	IGADTOYQDQIAVLRVNVSTGSIISVTSKTIISVEVQAGVRFRIIDYORLLGE				
	IGYTSNRPVMLCLMSAIEYGVIFLINDGIDRLMDLQMKADRLNDILVKYBELSVPE				
	S"				
	complement(1221..1637)				
	/gene="csgF"				
	complement(1221..1637)				
	/gene="csgF"				
	/codon_start=1				
	/transl_table=11				
	/product="Csgf protein"				
	/protein_id="CAA05313.1"				
	/db_xref="GI:2739235"				
	/db_xref="SWISS-PROT:O54292"				
	/translation="MRVGVAVVLLMLFSLPTWAGNMTFQFRNPNFGNPNNGSFLNS				
	AOANSYKDPAYNDPFIETPSALDNFTQALQSOGLLTINNGKRGKRVMTNDFII				
	DIARDQLOLVNTDRKTRGSTIEVSLQTSDF"				
	complement(1664..2059)				
	/gene="csgE"				
	complement(1664..2059)				
	/gene="csgE"				
	/codon_start=1				
	/transl_table=11				
	/product="CsgE protein"				
	/protein_id="CAA05314.1"				
	/db_xref="GI:2739236"				
	/db_xref="SWISS-PROT:O54293"				
	/translation="MKRYLTWIVAAELIFATGNLHANEVEVEVGLTLDHTVSSIGHE				
	FYRPSDKMESEYTGNTLINRSPARKMSMTITVNDVIFQTFLPFKRDEKTVVF				
	ALADTEALNRQIDQTLSTSLARDEF"				
	complement(2064..2714)				
	/gene="csgD"				
	complement(2064..2714)				
	/gene="csgD"				
	/function="transcriptional regulator"				
	/codon_start=1				
	/transl_table=11				
	/product="Csgd protein"				
	/protein_id="CAA05315.1"				
	/db_xref="GI:2739237"				
	/db_xref="GOA:O54294"				
	/db_xref="SWISS-PROT:O54294"				
	/translation="MFNEVSSHGHTLLITKPSLQATALLQHLKQSLATITKGLHNIG				
	RLSELDISAGCIVLMDMMEDAKKLIHYQDNLRSKRNNTKTLTLPDPYPRVLEIENP				
	HINGVFATDEQEHVSGLOGLIRGECFSGKSLASYLITHGNTRYNTEBSALLTHRS				
	KEINKLRIGASNNIARSLPISNTYKTHLYNLPKIAVAGNRQAVSWANDNLR"				
	complement(2869..2896)				
	/gene="csgDRC"				
	complement(2869..2896)				
	/gene="csgDRC"				
	complement(2891..2896)				
	/gene="csgDRC"				

```

gene
3344..3376
/gene="csGBA"
-35_signal
3344..3349
/gene="csGBA"
-10_signal
3371..3376
/gene="csGBA"
gene
3469..3924
/gene="csGB"
CDS
3469..3924
/gene="csGB"
/function="nucleator"
/codon_start=1
/evidence=experimental
/product="CsgB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/db_xref="SWISS-PROT:P55226"
/translation="MKNLKLPMLTITIGAPGATATYDARSEYFANVELSKSFPN
QAIIQGVGTDSARVROGSKLSVISOEGNNRRAVDAGNYNFAVLEQTNANDA
SISQAYGNSAIIQKSGSNANTTQGTQKTAVVVKQSHMAIRVQR"
3469..3531
/gene="csGB"
3532..3921
/gene="csGB"
/product="CsgB protein"
/function="nucleator"
3966..4421
/gene="csGB"
3966..4421
/gene="csGB"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/translation="MKLVAAFAAIVVGSALAGVVPOMGGGNGGNSGSPDST
LSIYQGSANAAALQSDARKSETTITGSGNGADVGCGDNSTIELTQNGFNNAT
IDWNAKNSDITVQGYGNNAALVNGTASDSVWVRQVGFNNATANOY"
3966..4025
/gene="csGA"
4026..4418
/gene="csGA"
/product="CsgA protein"
/function="fiber subunit"
4483..4809
/gene="csGC"
4483..4809
/gene="csGC"
/codon_start=1
/translation="MHTLILAIASNTITTTTQGDITTVIPQVTLNPPCVOYQIL
SVRDVGQSHQKQKLTSLPANQPIELSLVSNISSEDSVKIITVSDGSLHSQQ
WPSAQ"
4483..4533
/gene="csGC"
4534..4806
/gene="csGC"
/product="CsgC"
4940..5103
/gene="ORF103"
4940..5103
/gene="ORF103"
/codon_start=1

sig_peptide
mat_peptide
gene
CDS

```

ORIGIN

Query Match 87.7%; Score 400; DB 1; Length 5103;
 Best Local Similarity 92.3%; Pred. No. 1,4e-101;
 Matches 421; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCAATTCGACGATCGTAGTTTCGAGTGTGCT 60
DB 3966 ATGAACCTTTAAAGTGGAGCAATTCGACGATCGTAGTTTCGAGTGTGCT 60
QY 61 GCGCTCGTCCAAATGGGGGCGCGCGGTATCATTAACGGCGGCAATAGTCCGAC 120
DB 4026 GCGCTCGTCCAAATGGGGGCGCGCGGTATCATTAACGGCGGCAATAGTCCGAC 120
QY 121 CCGACTATGATGAGCTGTGTTACCCGTTGTTTACCCATGAATGCGACATGACTGCA 180
DB 4086 CCGACTATGAGCTGTGTTACCCGTTGTTTACCCATGAATGCGACATGACTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACATTAACCCAGCGGTTATGTATGAGCGCGAT 240
DB 4146 AGCGATGCCGCTAAATCTGAAACGACATTAACCCAGCGGTTATGTATGAGCGCGAT 240
QY 241 GTAGCCGAGGTGCGGATATAGTATGTAATGTAAGTCACTGAGTAATGTTTCAGAAAT 300
DB 4206 GTAGCCGAGGTGCGGATATAGTATGTAATGTAAGTCACTGAGTAATGTTTCAGAAAT 300
QY 301 GCCACCATGACGACATGGAAGCGTAAATCTCCATTAATCTGTGCGCAATACGGCGGT 360
DB 4266 GCCACCATGACGACATGGAAGCGTAAATCTCCATTAATCTGTGCGCAATACGGCGGT 360
QY 361 AATTAACGCGCGCTGTTAATCAAGCGCATCTGATTCACGCGTAATGTCGTCAGTT 420
DB 4326 AATTAACGCGCGCTGTTAATCAAGCGCATCTGATTCACGCGTAATGTCGTCAGTT 420
QY 421 GCTTTTGCAACAACGCGCAACGCTTAACCATATTA 456
DB 4366 GCTTTTGCAACAACGCGCTTAACCATATTA 4421

```

RESULT 4

AEO08749

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

CONSTRM

TITLE

JOURNAL

22411 bp DNA linear BCT 23-APR-2003
 AEO08749 AEO06468
 AEO08749.1 GI:16419641
Salmonella typhimurium LT2
Salmonella typhimurium LT2
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; *Salmonella*.
 1 (bases 1 to 22411)
 McCllland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
 Houtz, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
 Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
 Stoneking, T., Nhan, M., Waterston, R., and Wilson, R.K.
 Complete genome sequence of *Salmonella enterica* serovar *Typhimurium*
 LT2
 Nature 413 (6858), 852-856 (2001)
 2 (bases 1 to 22411)
 The *Salmonella typhimurium* Genome Sequencing Project
 Direct Submission
 Submitted (23-MAR-2001) Genome Sequencing Center, Department of

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
COMMENT Supported by NIH grant 5U 01 AI33283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs: GLIMMER, <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at ECoCyc; <http://ecocyc.pangeasystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguionB; http://kinch.cifn.unam.mx:8850/db/regulondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

FEATURES

SOURCE

1.22411
location/Qualifiers

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/note="LT2"

/gene="LT2"

/note="phoH"

/note="synonym: STM1126"

434. .439

/gene="phoH"

/note="putative RBS for phoH; RegulonDB:STMSIH001398"

454. .1308

/gene="phoH"

/note="similar to E. coli phoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="phoB-dependent ATP-binding pho regulon component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRVQENAGDVYLQKLVSMGRQKAVIKAREAKRVLRDS

BSHKQREESVTSIQMGVGAIGMARSDRTSPVKARNEQAHYLAINDSKQLFAT

GEAGCKTWISAKAALAIHKDVERIIVTRPVLDADGLPGDIAEKAPYRPV

YDVLKRTGASPMQYCLRPEIGKVIAPAVNRGRTPENAVYILDEAQNAAQKMF

LTRIGENTVIVNGDITQCCLPRGVRSGLSDALEKFEEDENVGIVHFKDCCVRSALC

ORTLHAYS"

complement(1414. .2303)

/gene="STM1127"

complement(1414. .2295)

/gene="STM1127"

/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 29%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcriptional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="MEQPPRLKPKKILDTGANKQKSTRAQRIAYITLAPPROVTO

SSIALSRETQAGEATVIRFCTLGKGFQDFKDLAIELATTESDSSPLDDEVSE

SDDAHAIQKONTISNVLSFTLLMDQOVYGVVDALRHGHSVYIPGVGSGITLAD

MKHKLRIQLRGDAVSNHFMVMTLLKAGDVAMGVSHSGTSPETVHSLRLAQGA

RBS

TTVATHNIGSPICEADPCLINGNRQGLQODSIGTKAQLFVFDLLYTLLVOSSPE
QARBSKLRPMALDMWTK"

complement(2298. .2303)

/gene="STM1127"

/note="putative RBS for STM1127; RegulonDB:STMSIH001399"

complement(2580. .4088)

/gene="STM1128"

complement(2580. .4076)

/gene="STM1128"

/note="similar to E. coli putative cotransporter (AAC76702.1); Blastp hit to AAC76702.1 (571 aa), 22%

identity in aa 7 - 478"

/codon_start=1

/transl_table=1

/product="putative sodium/glucose cotransporter"

/protein_id="AAL20059.1"

/db_xref="GI:16419644"

/translation="MITHSPGIVNVLVFGYLLAMLVGVFSRRQKTADVDFRGGR

VPGMAVGFATTLSSITFMSIPAKFTSDMTPIQGYLAIALIPLVFVFIPEPK

LKMSAVYELARPDPVRCRLPASMSPMLFHTGRALITPLVLAARPAIDPVLYL

LISVNCITTMGGIGEVITVDVIOGLLSSALIFIVITLKTQGGIDEFTVTOA

DKFPFATQFHSWSTETVPLVMIGLGFANIQFTASQVQVRYITDSIEETKTLT

NAKLVAVIPVEFPFAGSALFVYQHPQLPAGFNTGILPLFVTEMPVGIAGLIA

ALPAAQSSISSNSISSCNSDIYORLSHKRTPEPRMKIATKLIVLWGISAS

VMLVMADESEIMDAFNSLIGMGGMPTGLFMLGIFPKRANGSAYIGIISVITVIGA

RYADINPFYGVGSLSVTSCVIFAPLFPAPPTLDERPERVTL"

complement(4083. .4088)

/gene="STM1128"

/note="putative RBS for STM1128; RegulonDB:STMSIH001400"

complement(4413. .5099)

/gene="STM1129"

complement(4413. .5093)

/gene="STM1129"

/note="similar to E. coli putative enzyme (AAC76255.1);

Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 -

226"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20060.1"

/db_xref="GI:16419645"

/translation="MSLLARLEQGVHENGGLVSCQPVSPMDKPEIIVAAQAQAS

AGAVAVREGENRTVRLPHLSVPIIGIKSDTGSPPVRIPOVDVLAQAGDIT

AFDASFSRPPVDISILTRIRHGLMAADSTYNEGSKOKIEFTICVCTGTC

ITVYPPDLAVTQLSHAGCRVYAGRYNTPLAANALBHGMAVTVGSATIRIEHIQ

WFSHAVKX"

complement(5094. .5099)

/gene="STM1129"

/note="putative RBS for STM1129; RegulonDB:STMSIH001401"

5599. 6759

/gene="STM1130"

5599. 6759

/gene="STM1130"

/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

identity in aa 40 - 403"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20061.1"

/db_xref="GI:16419646"

/translation="MGOMKPKKQMTLMALCFVAITTSQVATTLDPIDPELKNQTS

AIDNGVTVGLGRTGTSWKRIDLKKQKBERKSPGQARBQSVSFLNDELVYRG

GVGRKNSBPLOVTSVDVYKTSPVKNATQKVTISPVGLTGHTGVKLNSTVLLITGVN

EHIPDKYFIDIAAADADESEKVKYIVYFNKPADVYFNKXVFIYNKAKENTKNAQGL

PGAGTAGSSVYGNPLMLINGELPGIARFTVYIIRAMVDNKLFWLKNSSQLPSPGQ

QOEGIAQFSGVSHGVYLVGGANFPAGKQVTKYKSHGKIKKKRDEYVGLINGH

MOVWGXKROPAGVSVSYGDBVFLIGENAKKQKPSVYSFTMRDGNLTIK"

6793. 7497

/gene="STM1131"

6793. 7497

/gene="STM1131"

/note="putative RBS for STM1131; RegulonDB:STMSIH001402"

CDS

6805..7497
 /gene="STM1131"
 /note="similar to E. coli orf, hypothetical protein
 (AAC77267.1); Blastp hit to AAC77267.1 (241 aa), 268
 identity in aa 4 - 239"
 /codon_start=1
 /transl_table=11
 /product="putative outer membrane protein"
 /protein_id="AAL20062.1"
 /db_xref="GI:16419647"
 /translation="MKINKYLLGMVSIAPSSYLQAAATLDYRHYADRTIRINDRAI
 IEKPNIGIFVDASKSGVDGDDKHLSDLVANALIELGYSINYKTDFVLOPEFI
 EESGPDITIKPKYLRGQYNDPSGVYMGGRYYPDAKTATNYSDEKTRPDYIGVFI
 DELKLEYNFTMDDSDQIKFDKNTKYNIEHNVALMKLNKSFYVEGVNVAVRNNITDER
 OTRVGVQVHPH"
 7768..9060
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7768..7773
 /gene="STM1132"
 /note="putative RBS for STM1132; RegulonDB:STMS1H001403"
 7780..9060
 /gene="STM1132"
 /note="similar to E. coli putative transport protein
 (AAC77235.1); Blastp hit to AAC77235.1 (425 aa), 578
 identity in aa 25 - 418"
 /codon_start=1
 /transl_table=11
 /product="putative sugar transport protein"

Gene

RBS

CDS

Query Match 87.7%; Score 400; DB 1; Length 22411;
 Best Local Similarity 92.3%; Pred. No. 1,6e-101;
 Matches 421; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGCGAGCATTCGCGACATGCTGATGCTCGGAGTCTCTGCT 60
 DB 17769 ATGAACCTTTAAAGTGCGAGCATTCGCGACATGCTGATGCTCGGAGTCTCTGCT 17828
 QY 61 GGGGTGCTTCCAAATGGCGCGCGGCGGTAATCATTAACGGCGCGCAATAGTCCGCG 120
 DB 17829 GGGGTGCTTCCAAATGGCGCGCGGCGGTAATCATTAACGGCGCGCAATAGTCCGCG 17888
 QY 121 CCGGACTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 180
 DB 17889 CCGGACTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 17948
 QY 181 AGGATGCGCGGTAATCTGAAACGACATTACCGAGCGGTTATGTTACGGCGCGAT 240
 DB 17949 AGGATGCGCGGTAATCTGAAACGACATTACCGAGCGGTTATGTTACGGCGCGAT 18008
 QY 241 GTAGGCGAGGCTGCGGATTAATGACTTGAATGACTGACTGATGCTTCAAGATTAAT 300
 DB 18009 GTAGGCGAGGCTGCGGATTAATGACTTGAATGACTGACTGATGCTTCAAGATTAAT 18068
 QY 301 GCCACATTCGACGATGGAACGCTTAAAACTCCGATATTAATCTCGGCCAATACGGCGGT 360
 DB 18069 GCCACATTCGACGATGGAACGCTTAAAACTCCGATATTAATCTCGGCCAATACGGCGGT 18128
 QY 361 AATAACGCGCGGCTGCTTAAATCAAGCGGATTCGACGCTGATGCTGCTGCTGCTGCT 420
 DB 18129 AATAACGCGCGGCTGCTTAAATCAAGCGGATTCGACGCTGATGCTGCTGCTGCTGCT 18168
 QY 421 GCTTTTGCAACAACGCCACGCGCTTACCAAGATTTAA 456
 DB 18189 GCTTTTGCAACAACGCCACGCGCTTACCAAGATTTAA 18224

RESULT 5
 LOCUS AL627269 254050 bp DNA linear BCT 04-JUL-2003
 DEFINITION *Salmonella enterica* serovar Typhi (*Salmonella typhi*) strain CT18,
 complete chromosome, segment 5/20.
 ACCESSION AL627269 AL513382
 VERSION AL627269.1 GI:16502231

KEYWORDS

SALMONELLA enterica subsp. enterica serovar Typhi

SOURCE

Salmonella enterica subsp. enterica serovar Typhi

ORGANISM

Bacteria: Proteobacteria; Gammaproteobacteria; Enterobacteriales;

REFERENCE

1 (bases 1 to 254050)

AUTHORS

Parkhill, J., Dougan, G., James, K.D., Thomson, N.R., Pickard, D.,

TITLE

Complete genome sequence of a multiple drug resistant *Salmonella*

JOURNAL

Nature 413 (6858), 848-852 (2001)

MEDLINE

21534947

PUBMED

11677608

REFERENCE

2 (bases 1 to 254050)

AUTHORS

Parkhill, J.

TITLE

Direct Submission

JOURNAL

Submitted (25-OCT-2001) Submitted on behalf of the *Salmonella*

COMMENT

Sequencing team, Sanger Centre, Wellcome Trust Genome Campus,

FEATURES

Location/Qualifiers

source

1..254050

gene

/organism="Salmonella enterica subsp. enterica serovar

CDS

/mol_type="genomic DNA"

misc_feature

/db_xref="taxon:90370"

misc_feature

/note="synonym: pepN"

misc_feature

/gene="STY1078"

misc_feature

/note="orthologue of E. coli pepN (AMPN ECOLI); Pasta hit

misc_feature

/product="aminopeptidase N"

misc_feature

/protein_id="CAD08183.1"

misc_feature

/db_xref="GOA:Q82770"

misc_feature

/db_xref="SPTREMBL:Q82770"

misc_feature

/translation="MTQOQAKYRHDYRQVQITDIEFLDDEKTVTAISOAVR

misc_feature

HGAPDAPLDGDLTVLSIHVNDAPWYKEEGALLISDPERPTLRIVEISPA

misc_feature

NTALEGLYQSGDALCTQCEAEGRHITWLDPRDVARPTTIDKSKYPLISNG

misc_feature

RVAQGELEENRHWOMODPPKPCYFAIVADDFVLRPTFTTSGREVALSLYDVG

misc_feature

NLRAPAMTSLKNSMKMDPTRGLEYDIDITVAVDFFNMGAEENGLNPNKYV

misc_feature

LARTDTRDQDYLDIRVIGHYTHWNTGRYTCDFQLSKDELTVFRDSESSDL

misc_feature

GSRAPARISVTRWRGLQFAEDASPMVARIKDYIENNNFYTLTVYEGAEVIMH

misc_feature

TLIGSENFQKQMOQLYERHDGSAATCDPVOQAMEASVNDLSHFRWYSGSTPIVY

misc_feature

KDYNEPTBQTYLLTISQRTPATDQAEKQPLHI PALEIELVDGKVIITLQGGHVA

misc_feature

LVNVTQABQTFEDNYFQVPAALCEFSAPVKELEKMSDOOLPTIMHANDBSRND

misc_feature

AAQSLATYIKLVNVAHQGQPLSLPVYVADAFRVILDEKIDPLALAEILTLPSAN

misc_feature

IAELFEVIDPIAIAQVREALTPTLAELADEFLATYNNHLDKRYVDHGDIGKTLRN

misc_feature

ACLRPLTFRGETLANTIVSKQYRDANNNTDALALUAAVAAQDLPORDTLMOEYDKRH

misc_feature

OGELVMDKRFIIQSTSPANVLETVRSLLKGRSFMSNPNIRSLIGFAAGSNPAFH

misc_feature

AQDGSQYQPLVEMLTDLNSRNQVARSRLIEPLIRKRYDDKROEMRAALBQLGLEN

misc_feature

LSGDLKTKYLA"

misc_feature

532..1332

misc_feature

/gene="STY1078"

misc_feature

/note="Pfam match to entry PF01433 Peptidase_M1, Peptidase

misc_feature

family M1, score 245.10, E-value 1.9e-84"

misc_feature

/gene="STY1078"

misc_feature

1060..1089

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

gene
RAINLGTATPMDLSILOSAGCLRAKLEII"
complement(5597..6295)
/gene="flagh"
/locus_tag="t1740"
complement(5597..6295)
/gene="flagh"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKYALHVPYMALVATLTGCAWIPAKPIYOGATTAPQIPGV
PVANGSIFOSAPINYGQLEDRPRNIGTLTIVIGENVSASSKSSANRDKGT
SEGFPIVPRYLOGLEFGNSRADMEASGNSFNKGASNAWNTSGLTIVTDVLNNG
LHVGEKQIAINQGTGFIRFSGVNPRTISGNSVPSIQVADARIEYGVNGYINBAON
MGLORFPLNLSPM"
complement(6350..7132)
/gene="flagh"
/locus_tag="t1741"
complement(6350..7132)
/gene="flagh"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flagh"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="WISSIMAKTGLDQAQTMVDYIANLVSTNGEGRQAVEDL
LYOTIRQPGAOSSQOTLPSSGQIGTVRPAVERLHSGQNLSTQNNKDAVAKQGF
EQVLPDPSATYARDSPQVNDQGLVYAGQVQVQVPAITIPANALITIGRDVSVT
QGGQAPVQVQGLNLTFTMNDRGLEISGENLTETOSGGAPESTPGNLGAGLYQGY
VETSNNVAEELNMIIQVRAVEINSKAVSTTDQMLQKLTOL"
complement(7146..7901)
/gene="flagh"
/locus_tag="t1742"
complement(7146..7901)
/gene="flagh"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flagh"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="WDHAIYTMGAASQTLANQAVTASNLAASTPGFRAOLNLRV
PVDGLSLATRLVLTASTPGADMTQGLDYTSRPLDVALQDDMLVQAADAGEYTRN
GNIQVPTGOLTIQGHPIVIGEGPTIVEGSEITIAAGQISALNPGDPNNVAVGR
LKIYVAEGNEVORSDGDLFRLTAQAERGAVALADPSIRMSGVLREGSNVPEVAMT
DMIAARPEBMQKVTISYDENEGRANQLMSNS"
complement(7922..9133)
/gene="flagh"
/locus_tag="t1743"
complement(7922..9133)
/gene="flagh"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flagh"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFQAVSGLMAATNLDVIGNNANSATYFPGSGTASFDMPA
GSKVLGVKAVAGITQDFDTGTTNTGRGLDAVISONGFRLVDSNGSVFYSRNGQFKL

gene
complement(5597..6295)
/gene="flagh"
/locus_tag="t1740"
complement(5597..6295)
/gene="flagh"
/locus_tag="t1740"
/note="corresponds to STY1219 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MOKYALHVPYMALVATLTGCAWIPAKPIYOGATTAPQIPGV
PVANGSIFOSAPINYGQLEDRPRNIGTLTIVIGENVSASSKSSANRDKGT
SEGFPIVPRYLOGLEFGNSRADMEASGNSFNKGASNAWNTSGLTIVTDVLNNG
LHVGEKQIAINQGTGFIRFSGVNPRTISGNSVPSIQVADARIEYGVNGYINBAON
MGLORFPLNLSPM"
complement(6350..7132)
/gene="flagh"
/locus_tag="t1741"
complement(6350..7132)
/gene="flagh"
/locus_tag="t1741"
/note="distal rod protein; corresponds to STY1218 from
Accession AL513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flagh"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="WISSIMAKTGLDQAQTMVDYIANLVSTNGEGRQAVEDL
LYOTIRQPGAOSSQOTLPSSGQIGTVRPAVERLHSGQNLSTQNNKDAVAKQGF
EQVLPDPSATYARDSPQVNDQGLVYAGQVQVQVPAITIPANALITIGRDVSVT
QGGQAPVQVQGLNLTFTMNDRGLEISGENLTETOSGGAPESTPGNLGAGLYQGY
VETSNNVAEELNMIIQVRAVEINSKAVSTTDQMLQKLTOL"
complement(7146..7901)
/gene="flagh"
/locus_tag="t1742"
complement(7146..7901)
/gene="flagh"
/locus_tag="t1742"
/note="corresponds to STY1217 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flagh"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="WDHAIYTMGAASQTLANQAVTASNLAASTPGFRAOLNLRV
PVDGLSLATRLVLTASTPGADMTQGLDYTSRPLDVALQDDMLVQAADAGEYTRN
GNIQVPTGOLTIQGHPIVIGEGPTIVEGSEITIAAGQISALNPGDPNNVAVGR
LKIYVAEGNEVORSDGDLFRLTAQAERGAVALADPSIRMSGVLREGSNVPEVAMT
DMIAARPEBMQKVTISYDENEGRANQLMSNS"
complement(7922..9133)
/gene="flagh"
/locus_tag="t1743"
complement(7922..9133)
/gene="flagh"
/locus_tag="t1743"
/note="corresponds to STY1216 from Accession AL513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flagh"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSFQAVSGLMAATNLDVIGNNANSATYFPGSGTASFDMPA
GSKVLGVKAVAGITQDFDTGTTNTGRGLDAVISONGFRLVDSNGSVFYSRNGQFKL

Query Match 87.4%; Score 398.4; DB 1; Length 301983;
Best Local Similarity 92.1%; Pred. No. 5.3e-101;

Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATCGTATGTTCTGGCAGTCTCTGCT 60
DB 37310 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATCGTATGTTCTGGCAGTCTCTGCT 37251
QY 61 GGGCGGTTCACATATGGGGGGGGGGGGTATATCAATACGGGGGGGCAATGTTCCGGC 120
DB 37250 GGGCGGTTCACATATGGGGGGGGGGGGTATATCAATACGGGGGGGCAATGTTCCGGC 37191
QY 121 CCGGACTATGATCAGCTGTTTACCCGTGTGTATCCCATGAAATGACATCAGTCA 180
DB 37190 CCGGACTATGATCAGCTGTTTACCCGTGTGTATCCCATGAAATGACATCAGTCA 37131
QY 181 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCCCGAT 240
DB 37130 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAACGGCCCGAT 37071
QY 241 GTAGGCCAGGGTGGCGGATATATGATCTATGTAAGTCACTCAAGAAAGTTTCAAAATAT 300
DB 37070 GTAGGCCAGGGTGGCGGATATATGATCTATGTAAGTCACTCAAGAAAGTTTCAAAATAT 37011
QY 301 GCCACCATCGACAGTGGAGCGCTAAAACTCCGATTAATCTGTCGCCCAATACGGCGGT 360
DB 37010 GCCACCATCGACAGTGGAGCGCTAAAACTCCGATTAATCTGTCGCCCAATACGGCGGT 36951
QY 361 AATAACGCCGCGCTGTTAATCAGCCGATCTGATTCGAGGTATGCTGATGCTGATGCT 420
DB 36950 AATAACGCCGCGCTGTTAATCAGCCGATCTGATTCGAGGTATGCTGATGCTGATGCT 36891
QY 421 GGTTCGCAACACGCGCAGCGCTACAGATTAA 456
DB 36890 GGTTCGCAACACGCGCAGCGCTACAGATTAA 36855
RESULT 7
STAGFBA
LOCUS STAGFBA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AJ000514
VERSION AJ000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE Salmomella typhimurium
ORGANISM Salmomella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J. and Rhen,M.
Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE
2 (bases 1 to 1048)
Sukupolvi,S.S.
Direct Submission
Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kivimyllykatu, 20520, FINLAND
FEATURES
source
1..1048
location/Qualifiers
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="agfB protein"
/protein_id="CAA04150.1"

```

/db_xref="GI:2275120"
/db_xref="SPTREMBL:O33801"
/translation="MKNLFLPMLTLTIGEPATATNYDLARSEYNFAVNELSKSPNO
AAILGQVGTDSARVRHSGSKLTVISOERGNRRARVDAQANFNPAITCORNADA
SISQAVGNASAIISALITOKSGNKNANITQGTAKTAVVVOKOSHAIQANITQYGTQ
KTAVVVOKOSHAIIRVTR"
593..1048
/gene="agfa"
593..1048
/gene="agfa"
/codon_start=1
/translation="agfa"
/product="agfa protein"
/protein_id="CAA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:O33802"
/translation="MKLLKVAFAIIVSGSAVAVVPQWGGGNGNGNSGPDST
LSIYQVSNAAALALQSDARKSETTITQSYGNADVGADVGADNISTELTONGFRNNAT
IDWNAKNSDITVQYGNNAALVNOTASDSSVMVROVFGNNAFANQYN"

```

ORIGIN

```

Query Match      86.1%; Score 392.6; DB 1; Length 1048;
Best Local Similarity 91.4%; Pred. No. 1.6e-99;
Matches 416; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

```

```

QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGCAATCGTATGTTCTGGCAGTCTTGCT 60
DB 593 ATGAACCTTTTAAAGTGGCAGCATTCGCAATCGTATGTTCTGGCAGTCTTGCT 652
QY 61 GGGGTGTTCCAAATGGGCGCGCGGATCATTAACGGCGCGCAATAGTTCCGCG 120
DB 653 GGGGTGTTCCAAATGGGCGCGCGGATCATTAACGGCGCGCAATAGTTCCGCG 712
QY 121 CCGGACTATGATCAGCTGTGTTACCGGTGTGTACCATGAATGGCAATGATCATG 180
DB 713 CCGGACTATGATCAGCTGTGTTACCGGTGTGTACCATGAATGGCAATGATCATG 772
QY 181 AGCGATGCCGTTAAATCTGAACGACCATTAACCCAGAGCGTTATGTAACGGCGCAT 240
DB 773 AGCGATGCCGTTAAATCTGAACGACCATTAACCCAGAGCGTTATGTAACGGCGCAT 832
QY 241 GTAGGCCAGGTCGGATATATGTAATGTAACGATCGTAACGATGTTTCAAAATAT 300
DB 833 GTAGGCCAGGTCGGATATATGTAATGTAACGATCGTAACGATGTTTCAAAATAT 892
QY 301 GGCACATGACCAAGTGGAACCTTAAACTCCGATTTACTGTGCGCCATTAACGGCG 360
DB 893 GGCACATGACCAAGTGGAACCTTAAACTCCGATTTACTGTGCGCCATTAACGGCG 952
QY 361 AATAAGCGCGCTGTTAATCAGACCGCATCTGATTCACCGTAATGTCGTCAGGTT 420
DB 953 AATAAGCGCGCTGTTAATCAGACCGCATCTGATTCACCGTAATGTCGTCAGGTT 1012
QY 421 GGTTCGCAACACGCGCGCTAACCAAGTATTA 455
DB 1013 GGTTCGCAACACGCGCGCTAACCAAGTATTA 1047

```

RESULT 8

```

LOCUS      144908      361 bp      DNA      linear      PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION  144908
VERSION    144908.1  GI:2469621
KEYWORDS
SOURCE
ORGANISM   Unknown.
REFERENCE   1 (bases 1 to 361)
AUTHORS    Doran, J. L., Kay, W. W., Collinson, S. Karan, and Clouthier, S. C.
TITLE      Methods and compositions comprising the agfa gene for detection of
JOURNAL    Patent: US 5635617-A 56 03-JUN-1997;

```

JOURNAL

```

FEATURES
source
Location/Qualifiers
1..361
/organism="unknown"
/mol_type="unassigned DNA"

```

ORIGIN

```

Query Match      61.9%; Score 282.2; DB 6; Length 361;
Best Local Similarity 90.1%; Pred. No. 1.9e-68;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```

```

QY 64 GTCTTCACAAATGGGCGGCGGATTAATCATTAACGGCGGCAATAGTCCGCCG 123
DB 1 GTCTTCACAAATGGGCGGCGGATTAATCATTAACGGCGGCAATAGTCCGCCG 60
QY 124 GACTATATCAGCTGTTTACCGGTGTTTACCATGAATGGCACATGCACTGCAAGC 183
DB 61 GACTATATCAGCTGTTTACCGGTGTTTACCATGAATGGCACATGCACTGCAAGC 120
QY 184 GATCCCGTAAATCTGAAGACGACATTAACCGAGGGTTATGTAACGGCCGATGTA 243
DB 121 GATCCCGTAAATCTGAAGACGACATTAACCGAGGGTTATGTAACGGCCGATGTA 180
QY 244 GGCAGGTCGGATTAATGATTAATGAACTGACTCAGAAATGTTTCAAAATATGCC 303
DB 181 GGCAGGTCGGATTAATGATTAATGAACTGACTCAGAAATGTTTCAAAATATGCC 240
QY 304 ACATGACACAGTGAAGCTAAACCTCCGATTAATGTTGCGCAATAGCGCGTAT 363
DB 241 ACATGACACAGTGAAGCTAAACCTCCGATTAATGTTGCGCAATAGCGCGTAT 300
QY 364 AACCGCGCGTGTATATCAGACCGCATGATTC 398
DB 301 AACCGCGCGTGTATATCAGACCGCATGATTC 335

```

RESULT 9

```

LOCUS      CSP515700      2889 bp      DNA      linear      BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csbB gene, csbA gene and csbD gene.
ACCESSION  AJ515700
VERSION    AJ515700.1  GI:31790491
KEYWORDS   csbA gene; csbB gene; csbD gene; curlin-csbA protein; nucleation
SOURCE     component of curlin monomers; regulatory protein.
ORGANISM   Citrobacter sp. Fec2
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Citrobacter.
REFERENCE   1
AUTHORS    Zogaj, X., Bokranz, W., Nimtz, M. and Romling, U.
TITLE      Production of Cellulose and Curli Fimbriae by Members of the Family
JOURNAL    Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
            Infect. Immun. 72 (7), 4151-4158 (2003)
REFERENCE   2 (bases 1 to 2889)
AUTHORS    Romling, U.
TITLE      Direct Submission
JOURNAL    Submitted (11-NOV-2002) Romling U., Microbiology and Tumorbiology
            Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
FEATURES
source
Location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csbD"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

FEATURES

```

source
Location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csbD"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

```

/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEVHSIHGHTLLITRPSLOATALLQHLKQSLALSGKLNIO
RSLDIISSCIVLVDMMEADKLIHYMDNLSRKNNILTKLLNPDDPYDIENMP
HNGVYVADDERVNVGLIGECYSQKLASYLITHSGNRYNSALLTHRE
KEILNKLRIGANIEIARSLFISNTVTKHLVNLFKIAVNRTOAVSANMNLNR"
1623..2078
/gene="cs8B"
1623..2078
/gene="cs8B"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKNKLIFMMLTVLGAFCIASATSYDLAHEYNFAVNELSKSPN
QALIGOVGTNNSAKRROGSKLISVSGSGNSAKDQSGAPAYIAQSGHSDA
SISQSNYGTAMIIQKSGSNKANITGYGKTAVVVQRQSMARIVQR"
2119..2571
/gene="cs8A"
2119..2571
/gene="cs8A"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-cs8A protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLOVAAPFAIIVSGSALAGVVPQWGGGGGGSSSGSPSTL
SIYQGVNNAALAIQSDARKSPDTTHONGFNGAGVGGSDNSTIDLRNGKKNATIT
DOMKNSDITVSYQGHNAALVNOTASDSVTVHQVFGNNAITANQ"

```

gene

CDS

Query Match 61.4%; Score 280; DB 1; Length 2889;
 Best Local Similarity 77.4%; Pred. No. 9e-68;
 Matches 353; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

```

OY 1 ATGAACCTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTCTGTGCT 60
DB 2119 ATGAACCTTTTCAAGTGGCAGCATTCGAGCAATCGTAGTTCTGGCAGTCTGTGCT 2178
OY 61 GCGGTCGTTCCAGATGGGGCGGCGCGGTATCATACGGGGCGGCATAGTCCGGC 120
DB 2179 GGTTCGTGTCGCAATGGGGCGGCGCGG--TGCGGGCGGGCGGAGCAGCTCCGCGC 2235
OY 121 CCGGACTATGATCAGCTGGTTACCCGCTGTGTTACCCATGAATGAGCAATGACATGCA 180
DB 2236 CCGGAAATGACCTTATGATTAATGATGACAGATGCAATGACCGCGCTTGCGTCA 2295
OY 181 AGCGATGCCGTAATCTGAAAGCAATTCACAGAGCGGTTATGTAACGGCGCCGAT 240
DB 2296 AGCGAGCTCGTAATCTGATACGACCAATTCAGAAATGCTTGTAACGGCGCAGAC 2355
OY 241 GTAAGCCAGGGTGGCGATTAATGTAATGTAAGTCAAGATGCTTTCGAATTAAT 300
DB 2356 GTGGGCGAGGGTCAAGATTAACGACCAATCGATTCGACTCAAAAGCGCTTCAAAAACAC 2415
OY 301 GCCACATCGACCAATGGAACGCTAAAACTCCGATTTATCTGCGGCAATATGCGCGGT 360
DB 2416 GCCACATCGATCAATGGAACGCAAAATTCGACATTTACTGTGAGCCAGTATGTGGA 2475
OY 361 AATAAGCCGCGCTGTTAATCAACCGCATGTAATTCAGCGTAATGTTGCGTCAAGTT 420
DB 2476 CATAAGCCGCGCTGTAACCAAGATGCGTCCGATTCAGAGGTTCTGTGTCATCAGGT 2535
OY 421 GGTTCGCAACAAGCGCAAGCTTAACAGATTAA 456
DB 2536 GGTTCGCAACAAGCGCAAGCTTAACAGATTAA 2571

```

RESULT 10
 LOCUS AF515701 2920 bp DNA linear BCT 24-JUN-2003
 DEFINITION Citrobacter freundii cs8B gene, cs8A gene and cs8D gene.
 ACCESSION AF515701
 VERSION AJ515701.1 GI:31790495
 KEYWORDS cs8A gene; cs8B gene; cs8D gene; curlin-cs8A protein; nucleation component of curlin monomers; regulatory protein.
 SOURCE
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Citrobacter.

REFERENCE
 AUTHORS Zogaj,X., Bokranz,W., Nimtz,M. and Romling,U.
 TITLE Production of Cellulose and Curli Fimbriae by Members of the Family Enterobacteriaceae Isolated from the Human Gastrointestinal Tract Infect. Immun. 72 (7), 4151-4158 (2003)
 JOURNAL 2 (bases 1 to 2920)
 TITLE Romling,U.
 JOURNAL Direct Submission
 Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source
 1..2920
 /organism="Citrobacter freundii"
 /mol_type="genomic DNA"
 /isolate="Fec4"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:546"
 /country="Germany"
 complement(211..861)
 /gene="cs8D"
 complement(211..861)
 /gene="cs8D"
 /function="necessary for transcription of the cs8A operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CAD56673.1"
 /db_xref="GI:31790496"
 /translation="MFNEVHSIHGHTLLITRPSLOATALLQHLKQSLALSGKLNIO
 RSLDIISSCIVLVDMMEADKLIHYMDNLSRKNNILTKLLNPDDPYDIENMP
 HNGVYVADDERVNVGLIGECYSQKLASYLITHSGNRYNSALLTHRE
 KEILNKLRIGANIEIARSLFISNTVTKHLVNLFKIAVNRTOAVSANMNLNR"
 1624..2082
 /gene="cs8B"
 1624..2082
 /gene="cs8B"
 /function="nucleator"
 /codon_start=1
 /transl_table=11
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56674.1"
 /db_xref="GI:31790497"
 /translation="MKNKLIFMMLTVLGAFCIASATSYDLAHEYNFAVNELSKSP
 NQALIGOVGTNNSAKRROGSKLISVSGSGNSAKDQSGAPAYIAQSGHSDA
 SISQSNYGTAMIIQKSGSNKANITGYGKTAVVVQRQSMARIVQR"
 2123..2572
 /gene="cs8A"
 2123..2572
 /gene="cs8A"
 /function="curli subunit"
 /codon_start=1
 /transl_table=11
 /product="curlin-cs8A protein"
 /protein_id="CAD56675.1"
 /db_xref="GI:31790498"
 /translation="MKLOVAAPFAIIVSGSALAGVVPQWGGGGGGSSSGSPSTL
 SIYQGVNNAALAIQSDARKSPDTTHONGFNGAGVGGSDNSTIDLRNGKKNATIT
 DOMKNSDITVSYQGHNAALVNOTASDSVTVHQVFGNNAITANQ"

ORIGIN

Query Match 53.8%; Score 245.2; DB 1; Length 2920;
 Best Local Similarity 72.8%; Pred. No. 5,8e-58;
 Matches 332; Conservative 0; Mismatches 118; Indels 6; Gaps 1;

QY 1 ATGAACTTTTAAAGGAGCATTCGAGCAATGCTGTTCTGGCAGTCTGCT 60
 DB 2123 ATGAACTTTTAAAGGAGCATTCGAGCAATGCTGTTCTGGCAGTCTGCT 2182
 QY 61 GGGTGTCTTCCACAAATGGGCGCGCGGTTATCATTAACGCGCGCGCAATAGTCCGGC 120
 DB 2183 GGGTGTCTTCCACAAATGGGCGCGCGGTTATCATTAACGCGCGCGCAATAGTCCGGC 2236
 QY 121 CCGAGTATGATCAGTGTGTTACCGGTGTGTTACCCATGAAATGGCAATGACATGCA 180
 DB 2237 CCGAGTATGATCAGTGTGTTACCGGTGTGTTACCCATGAAATGGCAATGACATGCA 2296
 QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGCGCGCGAT 240
 DB 2297 AGTGTATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTATGTAACGCGCGCGAT 2356
 QY 241 GTAGCCAGAGGCGGATTAATGTAATCTTGAATCTGACTGAAATGGTTTCAGAAATTAAT 300
 DB 2357 GTAGCCAGAGGCGGATTAATGTAATCTTGAATCTGACTGAAATGGTTTCAGAAATTAAT 2416
 QY 301 GCCACATCGAACAGTGGAAAGCTTAAATCCGATATTACTGTGCGCCCAATCGCGCGT 360
 DB 2417 GCCACATCGAACAGTGGAAAGCTTAAATCCGATATTACTGTGCGCCCAATCGCGCGT 2476
 QY 361 AATAAGCCGCGCGGTATTAATCAAGCCGATCTGATTCAGCGCTAATGTCGTCAGATT 420
 DB 2477 CGCAACGCGGTGTGTTAATCAAGCCGCTCTGACTTCCAAATGTCTGATTCAGCAGGTT 2536
 QY 421 GGTTTGGCAACAAAGCGCAGCGCTTAACCAATATTAA 456
 DB 2537 GGTTTGGTAAACAAAGCTTACAGCTTAACCAACTAA 2572

RESULT 11
 ECCSGABDG 4680 bp DNA linear BCT 07-JUL-2002
 LOCUS E.coli csqG, csqF, csqE, csqD, csqB, csqA, and orfC genes.
 DEFINITION X90754.1 GI:1147558
 ACCESSION csqA gene; csqB gene; csqD gene; csqE gene; csqF gene;
 VERSION orfC gene.
 KEYWORDS Escherichia coli;
 SOURCE Escherichia coli;
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1
 AUTHORS Hammar, M., Arnyqvist, A., Bian, Z., Olesen, A. and Normark, S.

TITLE Expression of two csq operons is required for production of

JOURNAL coli K-12 and congo red-binding curli polymers in Escherichia

MEDLINE Mol Microbiol. 18 (4), 661-670 (1995)

PUBMED 96414468

REFERENCE 8817489

AUTHORS 2 (bases 1 to 4680)

JOURNAL Hammar, M.

TITLE Direct Submission

Microbiology and Immunobiology Center, Box 280, S-171 77 Stockholm,

SWEDEN

FEATURES

SOURCE

1. 4680

/organism="Escherichia coli"

/mol_type="genomic DNA"

/strain="K12"

/db_xref="taxon:562"

/map="23.15 minutes"

1. 3140

/organism="Escherichia coli"

source

repeat_region

gene

CDS

gene

CDS

gene

CDS

gene

CDS

gene

CDS

misc_feature

/mol_type="genomic DNA"
 /strain="MC4100"
 /db_xref="taxon:562"
 3141..4680
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="WJ110"
 /db_xref="taxon:562"
 <1..>4680
 /insertion_seq="IS2 (partial)"
 complement(133..966)
 /gene="csqG"
 complement(133..966)
 /gene="csqG"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 /protein_id="CA62277.1"
 /db_xref="GI:1147559"
 /db_xref="GOA:P52103"
 /db_xref="SWISS-PROT:P52103"
 /translation="MORLFLVAVMLSGCLTAPRKEARPTLMPRAQSYDYLTHLPA
 PTGRIFVSYVNIODETGOFKPYPAENPSTAVPQSATLAVLAKDSRFILERQGLQ
 NLINERKIIRAAQENGTVA INNRILQSLTAVNVEBSIIIGESNVKSGGVGARVFG
 IGADROYODQAVNLRVNVSTGSLISVNTSKTILSEVQAGVFEDIDYRLLGE
 VGTSTNEPMLCLMSALETGVLEFLINDGIDRGLMDLQWKARQNDILVYKHHMSVPE
 S"
 complement(993..1409)
 /gene="csqF"
 complement(993..1409)
 /gene="csqF"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 /protein_id="CA62278.1"
 /db_xref="GI:1147560"
 /db_xref="SWISS-PROT:P52104"
 /translation="MRVHAVVILMLISPSMACTMTFQFRNPFQGNPNNGAFLNS
 AOAQNSYDPSYNDRCIEPTSLADNTQAIQSLIGGLSNINTKPGRMVNDYIV
 DIARDQLOLVNDRKGTGTSTIOVSGLONNSTDF"
 complement(11434..1823)
 /gene="csqE"
 complement(1434..1823)
 /gene="csqE"
 /codon_start=1
 /transl_table=11
 /product="assembly /transport component in curli
 /protein_id="CA62279.1"
 /db_xref="GI:1147561"
 /db_xref="SWISS-PROT:P52105"
 /translation="MKRYLWIVAAEFPLPAAGNLHAAVEVPGLLTDTWTSIGHDYF
 RASDWSQSDYNTLFTINRPSARMSWTTITVNDVIFQTFLEPKRDFEKTVAFL
 IOTEBALNROINQALSTGDAHDEF"
 complement(1828..2478)
 /gene="csqD"
 complement(1828..2478)
 /gene="csqD"
 /note="necessary for transcription of the csqA operon"
 /codon_start=1
 /transl_table=11
 /product="putative regulatory protein"
 /protein_id="CA62280.1"
 /db_xref="GI:1147562"
 /db_xref="GOA:P52106"
 /db_xref="SWISS-PROT:P52106"
 /translation="MFNEVSHIGHHTLLITKSLQATALLQHKSLATLTKLHNIQ
 RSLDSSGSIILLDMEDAKDLIHYMDPLSKNNNIKILTLNTPDYRDIENWP
 HINGVYEMEDQERVNGLQVARECTFQKASLYITISGVNRYNSTESALLTHRE
 KEINKRLIGASNNRIARSLFISNTVKTLYNLFKKIAYKNRTOAVSMANDMLRR"
 complement(2626)

```

misc_feature      /note="transcriptional start site for csdDEFG operon"
                  3140
gene              /note="transcriptional start site for csdGA operon"
                  3233..3688
                  /gene="csdG"
                  3233..3688
                  /gene="csdG"
                  /codon_start=1
                  /transl_table=11
                  /product="nucleation component of curlin monomers"
                  /protein_id="CAA62281.1"
                  /db_xref="GI:1147563"
                  /db_xref="GOA:P39828"
                  /db_xref="SWISS-PROT:P39828"
                  /translation="MKKLLFMKLLFGLTGAAGTAAAGVLANSEYFPAVNEKSSPN
OAAITGAGTNNSSAOLROGCKSLAVVAOEGSSNPAKIDQTDYINLAIYIDQKGSANDA
SISQGAAYNTAIIQKSGNKNKITQYGTAKTAVVORQSMARIVTOR"
                  3729..4184
gene              /gene="csdG"
                  3729..4184
                  /gene="csdG"
                  /note="major subunit"
                  /codon_start=1
                  /transl_table=11
                  /product="curlin/CsdG protein"
                  /protein_id="CAA62282.1"
                  /db_xref="GI:1147564"
                  /db_xref="GOA:P28307"
                  /db_xref="SWISS-PROT:P28307"
                  /translation="MKLLKVAIAIAIVFSGSLAVGPVQGGGNNHGGNNSGPNS
LNIYQGGNSALALQTDARNBDLITQHGCGNGADVGCGSDSIDLTQRFQNSAT
LDQWKNSEMVVKQFGGNGAAYVDTASNSVNTQVFGNNAATHQ"
                  3729..3788
sig_peptide       /gene="csdG"
                  4243..4575
gene              /gene="orfC"
                  4243..4575
                  /gene="orfC"
                  /codon_start=1
                  /transl_table=11
                  /product="putative curli production protein"
                  /protein_id="CAA62283.1"
                  /db_xref="GI:1147565"
                  /db_xref="SWISS-PROT:P52107"
                  /translation="MTMLILAAISSQITPNTQGGDYTTITPEVTLTQSCICRYOIL
SURESQSQTQKRTKLSLPANOPIALTKLSLNTSPDRKIVVTVDGOSLHLSQO
MPPSEKS"
ORIGIN
Query Match      50.9%; Score 232; DB 1; Length 4680;
Beet Local Similarity 69.3%; Pred. No. 3,1e-54;
Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;
QY 1 ATGAACCTTTTAAAGTGCGACGACATTCGAGCAATCGTATTCTGCGAGTCTGTGCT 60
    |||||
DB 3729 ATGAACCTTTTAAAGTGCGACGACATTCGAGCAATCGTATTCTGCGAGTCTGTGCGA 3788
    |||||
QY 61 GCGCGTGTTCACAAATGGGCGCGCGCGGTAAATCATACGCGCGGCGCAATAGTTCCGCG 120
    |||||
DB 3789 GGTGTGTTCCTCAGACGCGCGCGCGGTAAATCATACGCGCGGCGGTAAATAGCGCGC 3848
    |||||
QY 121 CCGGACTATGATGAGTGGTATCCGCGTGTGTTATCCCATGAATGCGCATGCACTGCA 180
    |||||
DB 3849 CCAAAATCTTACCTGAACATTTACAGTACGGGTGGGGAATCTGCACTTCTCTGCA 3908
    |||||
QY 181 ACCGATGCCCGTAAATCTGAAAGCAACATTACCCAGAGCGGTTATGTAAACGCCCGCAT 240
    |||||
DB 3909 ACTGATGCCCGTAACTCTGACTTGAATTAACCAAGCAAGGCGGCGGTAAATGTGCAAGT 3968
    |||||
QY 241 GTAGGCGAGGGTGGCGATATATGTAATTTGAATGACTGACAGATGGTTTCGAAATTAAT 300
    |||||
DB 3969 GTTGTGTAGGGCTCAGATGACGCTCAATGCAATGCAATGCAACGATGCTTCCGTAAACGC 4028

```

```

QY 301 GCCACATGACAGTGGAGACGCTAAAAATCTCGATATTAATGTCGCGCAATACGGCGGT 360
    |||||
DB 4029 GCTACTCTTGATGATGAGACGCGCAAAATTTGAAATGACGCTTAAACAGTTCCGTGTGT 4088
    |||||
QY 361 AATAACGCCCGCGCTGTTAATACAGCCGATCTGATTCAGCGTAAATGATGCGTACGTT 420
    |||||
DB 4089 GGCACGCGTGTGACGATTGACAGACGATCATTAATCTCTCCGTCACAGTACTACGTT 4148
    |||||
QY 421 GGTTTGGCACACACGACGCGCTAACGATTTAA 456
    |||||
DB 4149 GGTCTGTGTAACACGACGCGCTCATCATTA 4184
    |||||

RESULT 12
LOCUS      AE000205 10346 bp DNA linear BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 95 of 400 of the complete genome.
ACCESSION  AE000205 U00096
VERSION     AE000205.1 GI:1787265
KEYWORDS   SOURCE
            ORGANISM
            Escherichia coli K12
            Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
            Enterobacteriaceae; Escherichia.
            1 (bases 1 to 10346)
            Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
            Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
            Gregor, J.J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
            Mau, B. and Shao, Y.
            The complete genome sequence of Escherichia coli K-12
            Science 277 (5331), 1453-1474 (1997)
            97426617
            MEDLINE
            9278503
            PUBLISHED
            2 (bases 1 to 10346)
            Blattner, F.R.
            Direct Submission
            Submitted (16-JUN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            3 (bases 1 to 10346)
            Blattner, F.R.
            Direct Submission
            Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecol@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
            4 (bases 1 to 10346)
            Plunkett, G. III.
            Direct Submission
            Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 Strain MG1655. Predicted open reading
            frames were determined using Genemark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 (e-mail: markamber.gatech.edu). Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). ** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications

```

and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'p' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

SOURCE

Location/Qualifiers
1..10346
/organism="Escherichia coli K12"
/mol_type="genomic DNA"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
complement(<1..48)
/note="153"

repeat_region

promoter

/note="factor Sigma70; predicted +1 start at 1094689"
5..32
/note="factor Sigma70; predicted +1 start at 1094716"

gene

CDS

69..392
/gene="b1028"
/function="orf; Unknown"
/note="o107; This 107 aa ORF is 27 pct identical (0 gaps) to 58 residues of an approx. 1888 aa protein PAS2_CANAL SW: P43098"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74113.1"
/db_xref="GI:1787266"
/translation="MEDGYQAMRDVLAIGDIASPHQVLMVTKASEVAPSIITENUS LHTHTVKKTESLDTFLNPTVAFLEISNKPYLEGNMLNDFAGVDQQRVHCRAFSDAVPRK"
275..299
/gene="b1028"
/note="factor Sigma32; predicted +1 start at 1094983"
389..1375
/gene="ycdu"
/note="synonym: b1029"
389..1375
/gene="ycdu"
/function="orf; Unknown"
/note="o328; This 328 aa ORF is 19 pct identical (4 gaps) to 191 residues of an approx. 288 aa protein Y320_MYCSE SW: P47562"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74114.1"
/db_xref="GI:1787267"
/translation="MIPDYLFIRFODKRNLIYVATGLIIGPYWKNAGFTPSSEDI GVSGILALVYNFIPDLKAYMAKCYTKNIDSPFKKONHKELELTQPLVAGFLS LIMSAMSGCYOLPSPYALFISLIGPLVIFPMTIRSYKQVALSAVAKVYK SLRTYVLSICISTIVNDLITSPUNSDSVTEQOMLTFKSIITALLICGVVAIMLF FLRFSKYAFGLFQEIIDLFFSSENALSTFFAKPLMLFLILVEMVITLVSVL ATIVEMRIWEFAVFLCYVPLIYFFYCRFLMHNDFMACDMYPRMGHNK"

protein_bind

protein_bind

promoter

gene

CDS

/note="central position to predicted promoter: -104"
/bound_molety="Deor predicted site"
1253..1280
/gene="ycdu"
/note="No predicted promoter"
/bound_molety="Deor predicted site"
1337..1365
/gene="ycdu"
/note="factor Sigma70; predicted +1 start at 1096049"
1494..1745
/gene="b1030"
1494..1745

/gene="b1030"
/function="orf; Unknown"
/note="o83; This 83 aa ORF is 25 pct identical (7 gaps) to 81 residues of an approx. 616 aa protein ALBU_RABIT SW: P49065"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC74115.1"
/db_xref="GI:1787268"
/translation="MGGTRCVQYTKKPKVLSYELFPKYGGEGIDLSLSPCGPAHC VHGLSNLSPVVDPRSGILIPVCAIHEKARFVAL"
1513..1926
/gene="ycdu"
/note="synonym: b1031"
1513..1926
/gene="ycdu"
/function="putative structure; Not classified"
/note="o137; This 137 aa ORF is 27 pct identical (1 gap) to 47 residues of an approx. 80 aa protein R337_YEAST SW: P05759"
/codon_start=1
/transl_table=11
/product="putative ribosomal protein"
/protein_id="AAC74116.1"
/db_xref="GI:1787269"
/translation="MCNIRKSDPYRPTSSILMAVRGLTRPARAGSPITAFYTCPT GCRLSTPGRFSPPRCVOYTKKPKVLSYELFPKYGGEGIDLSLSPCGPAHC VHGL SNLSPVPRSGILIPVCAIHEKARFVAL"
complement(2111..2198)
/gene="sexX"
/note="synonym: b1032"
complement(2111..2198)
/gene="sexX"
/product="tRNA-Ser"
/note="anticodon: GGA, CG Site No. 17878"
/function="RNA; tRNA"
2269..2296
/note="factor Sigma70; predicted +1 start at 1096980"
2338..2365
/note="factor Sigma70; predicted +1 start at 1097049"
2355..2383
/note="factor Sigma70; predicted +1 start at 1097067"
2393..3370
/gene="ycdu"
/note="synonym: b1033"
2393..3370
/gene="ycdu"
/function="putative enzyme; Not classified"
/note="o325; This 325 aa ORF is 32 pct identical (2 gaps) to 177 residues of an approx. 312 aa protein YPRA_CORGL SW: P45637"
/codon_start=1
/transl_table=11
/product="putative dehydrogenase"
/protein_id="AAC74117.1"
/db_xref="GI:1787270"
/translation="WRINFAQRSGVSMIIFPHPTDQMWTEALRKAIPQAVRAM KSGNDSADVALVWHPPEMLAGDILAVPALGAVDSIISKLAQHEPMNPSVPLR LEDTGMQMOEYAVSOVLHFRFEDYRACQNSHWPQPEYHREDFTIGAGVYL GSKYASIQTRFPLRCMSRTRKSPGVQSFAGSELSARLSCRVINILPPTPV GIINOULEKLPDGAIVLNLARGVHVEDLLAALDSGKYGKMLVFNREPLPSPS LMOHPRVITIPHVAIIRPAEAVYISRTIAQLEKGRVCGQVDRARGI"

CDS

gene

gene

tRNA

promoter

promoter

promoter

gene

CDS

Query Match
Best Local Similarity 50.9%; Score 232; DB 1; Length 10346;
Matches 316; Conservativity 0; Mismatches 140; Indels 0; Gaps 0;
promoter
/note="factor Sigma70; predicted +1 start at 1097992"
3383..3410
/note="factor Sigma70; predicted +1 start at 1098094"
3396..3423

/organism="Escherichia coli"
 /mol_type="unassigned DNA"
 /db_xref="taxon:562"
 1..456
 /note="unassigned protein product"
 /codon_start=1
 /transl_table=11
 /protein_id="CAE83541.1"
 /db_xref="GI:39104002"
 /translation="MKLIKVEALIAIVFSGSALAGVPOYGGGNGHGGGNGNSGPNSE
 LNIYQGGNSALALQTDARNSDLITQHGNGADVGQSDSSIDLITQGFNSAT
 LDQWNGKNSMTVVKQFGGNGCAAVDQTAANSNVVTVQFGNNATAHQY"

ORIGIN

Query Match 50.5%; Score 230.4; DB 6; Length 456;
 Best Local Similarity 69.1%; Pred. No. 7.6e-54;
 Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGGAGTGTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGGAGTGTGCT 60
 QY 61 GCGCTGTTCCACAAATGGGCGCGCGGTATCATTAACGCGCGCGCAATAGTCCGAC 120
 DB 61 GGTGTGTTCTCAGTACGGCGCGCGGTATCATTAACGCGCGCGCAATAGTCCGAC 120
 QY 121 CCGGACTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCACATGCTGCA 180
 DB 121 CCAATTCGTGAGCTGAACATTTACAGTACGCTGCGGTACTCTGCACTTGTCTGCA 180
 QY 181 AGCGATGCCCGGTAATCTGAAACGACATTCACGAGCGGTTATGTAACGCGCGCAT 240
 DB 181 ACTGATGCCCGGTAATCTGAGCTTGTGACTTATCCGACATGCGCGCGGTATGTAACG 240
 QY 241 GTAGCCAGGCTGCGGATTAATGATCTATTGAACTGACTGCAATGTTTCAGAAATAT 300
 DB 241 GTTGTGAGGCTCAGATGACGCTCAATGATCTGACCAACGTTGCGGTATGACG 300
 QY 301 GCCACATCGACGATGGAAGCGCTTAAATCTCCGATTTACTGTGGCCATACGCGCGT 360
 DB 301 GCTACTCTTGATGATGGAACGCGCAAAATCTGAAATGACGCTTAAACAGTTCGCTGT 360
 QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 420
 DB 361 GGCACGCTGTGCTGACGATGACGACGCTGATCTCTCCGTCAACGATGACGAT 420
 QY 421 GGTTTGGCAACAAACGCGCGCTTACCAAGATTTAA 456
 DB 421 GGTTTGGTAAACAAACGCGCGCTTACCAAGATTTAA 456

RESULT 15

ECOCGGA 648 bp DNA linear BCT 13-JUL-1993
 DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
 VERSION L04979.1 GI:290424
 KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
 SOURCE Escherichia coli
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 648)
 Olesen, A., Arntqvist, A., Hammar, M., Sukupolvi, S. and Normark, S.
 The RpoS sigma factor relieves H-NS-mediated transcriptional
 repression of csgA, the subunit gene of fibronectin-binding curli
 in Escherichia coli

JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)
 MEDLINE 93311294
 PUBMED 8459772

COMMENT On Jun 11, 1993 this sequence version replaced gi:145630.
 Original source text: Escherichia coli (sub_strain W3110, strain
 K-12) (library: Kohara) DNA.

FEATURES

source

location/Qualifiers

1..648
 /organism="Escherichia coli"
 /mol_type="genomic DNA"
 /strain="K-12"
 /sub_strain="W3110"
 /db_xref="taxon:562"
 /rbsuse_1ib="Kohara"
 83..538
 /gene="csgA"
 83..538
 /gene="csgA"
 /note="precursor protein"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="curlin subunit"
 /protein_id="AAA23616.1"
 /db_xref="GI:290425"
 83..142
 /gene="csgA"
 /evidence=experimental
 143..535
 /gene="csgA"
 /product="curlin subunit"
 /evidence=experimental

CDS

gene

sig_peptide
 mat_peptide

ORIGIN

Query Match 50.5%; Score 230.4; DB 1; Length 648;
 Best Local Similarity 69.1%; Pred. No. 7.8e-54;
 Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGGAGTGTGCT 60
 DB 83 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCTGGGAGTGTGCT 60
 QY 61 GCGCTGTTCCACAAATGGGCGCGCGGTATCATTAACGCGCGCGCAATAGTCCGAC 120
 DB 143 GGTGTGTTCTCAGTACGGCGCGCGGTATCATTAACGCGCGCGCAATAGTCCGAC 120
 QY 121 CCGGACTATGATCAGCTGTTACCGGTGTTGTTACCCATGAAATGGCACATGCTGCA 180
 DB 203 CCAATTCGTGAGCTGAACATTTACAGTACGCTGCGGTACTCTGCACTTGTCTGCA 180
 QY 181 AGCGATGCCCGGTAATCTGAAACGACATTCACGAGCGGTTATGTAACGCGCGCAT 240
 DB 263 ACTGATGCCCGGTAATCTGAGCTTGTGACTTATCCGACATGCGCGCGGTATGTAACG 240
 QY 241 GTAGCCAGGCTGCGGATTAATGATCTATTGAACTGACTGCAATGTTTCAGAAATAT 300
 DB 323 GTTGTGAGGCTCAGATGACGCTCAATGATCTGACCAACGTTGCGGTATGACG 300
 QY 301 GCCACATCGACGATGGAAGCGCTTAAATCTCCGATTTACTGTGGCCATACGCGCGT 360
 DB 383 GCTACTCTTGATGATGGAACGCGCAAAATCTGAAATGACGCTTAAACAGTTCGCTGT 360
 QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTTGCGTCA 420
 DB 443 GGCACGCTGTGCTGACGATGACGACGCTGATCTCTCCGTCAACGATGACGAT 420
 QY 421 GGTTTGGCAACAAACGCGCGCTTACCAAGATTTAA 456
 DB 503 GGTTTGGTAAACAAACGCGCGCTTACCAAGATTTAA 456

Search completed: March 15, 2004, 22:50:13
 Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Scoring table: IDENTITY_NUC

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :
1:  _Genseqgn20190as:*
2:  _Genseqgn1990s:*
3:  _Genseqgn2000s:*
4:  _Genseqgn2001as:*
5:  _Genseqgn2001bs:*
6:  _Genseqgn2002s:*
7:  _Genseqgn2003as:*
8:  _Genseqgn2003bs:*
9:  _Genseqgn2003cs:*
10: _Genseqgn2004s:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	456	3	AAC64625	AAC64625 Agfa::PT3
2	406.4	89.1	456	2	AAQ87467	AAQ87467 Agfa sequ7
3	406.4	89.1	456	2	AAAT74142	AAAT74142 Salmonella
4	406.4	89.1	456	3	AAC64617	AAC64617 Salmonella
5	353.2	79.6	456	3	AAC64626	AAC64626 Agfa::PT3
6	353.2	77.9	456	3	AAC64628	AAC64628 Agfa::PT3
7	355.2	77.9	456	3	AAC64622	AAC64622 Agfa::PT3
8	355.2	77.9	456	3	AAC64629	AAC64629 Agfa::PT3
9	355.2	77.9	456	3	AAC64623	AAC64623 Agfa::PT3
10	353.6	77.5	456	3	AAC64624	AAC64624 Agfa::PT3
11	352	77.2	456	3	AAC64630	AAC64630 Agfa::PT3
12	352	77.2	456	3	AAC64627	AAC64627 Agfa::PT3
13	348.8	76.5	456	3	AAC64631	AAC64631 Agfa::PT3
14	282.2	61.9	361	2	AAQ73066	AAQ73066 Agfa sequ9
15	282.2	61.9	361	2	AAAT74141	AAAT74141 Salmonella
16	230.4	50.5	456	3	AAC64619	AAC64619 Escherichia
17	230.4	47.8	456	3	ACF36153	ACF36153 E. coli C
18	218	40.5	646	2	AAQ62647	AAQ62647 Fibronectin
19	156	34.2	369	2	AAQ62646	AAQ62646 FMB curli1 K
20	51.2	11.2	100	7	ACD68807	ACD68807 E. coli K
21	50	11.0	78	3	AAC64609	AAC64609 Agfa (SEF)
22	50	11.0	78	3	AAC64610	AAC64610 Agfa (SEF)
23	48.2	10.6	78	9	ACF36151	ACF36151 E. coli C

24	48	10.5	48	3	AA664621	AA66621	Leishmani
25	43.2	9.5	48	3	AC64416	AA66616	S. enterti
26	42.4	9.3	100	7	ACD68808	ACD68808	E. coli K
27	41.2	9.0	78	3	AC64606	AC64606	Seta (SEF
28	40.8	8.9	100	7	ACD68809	E. coli K	Seta (SEF
29	40.6	8.9	78	3	AA64605	AA64605	Seta (SEF
30	40.4	8.6	47475	3	AAAB1465	AAAB1465	N. mening
31	39.4	8.6	110000	3	AAAB1489	Continuation	(6 of
32	39.4	8.6	349900	3	AA21612	Continuation	(6 of
33	39.2	8.6	1003	5	AA576745	AA576745	DNA encod
34	37	8.1	14652	3	AAAB1482	AAAB1482	N. mening
35	37	8.1	110000	3	AAAB1489	Continuation	(2 of
36	37	8.1	349900	3	AA21611	Continuation	(2 of
37	36.4	8.0	2610	7	ACH42770	ACH42770	Prokaryot
38	36	7.9	2667	7	ACH4515	ACH4515	Prokaryot
39	35.8	7.9	954	7	ACA25910	ACA25910	Prokaryot
40	35.6	7.8	2000	7	ADA71938	ADA71938	Rice gene
41	35.4	7.8	3300	6	AB667377	AB667377	Neisseria
42	35.4	7.8	3411	5	AA588526	AA588526	DNA encod
43	35.4	7.8	3412	5	AA589144	AA589144	DNA encod
44	34.8	7.6	2000	7	ADA71938	ADA71938	Rice gene
45	34.6	7.6	456	3	AA64620	AA64620	Bacterich

ALIGNMENTS

RESULT 1
 AAC64625
 ID AAC64625 standard; DNA; 456 BP.
 XX
 AC AAC64625;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
 XX
 KW Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
 RV vaccine; immune response; immunogen; ds.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 XX
 PN MO20060102-A2.
 PD
 BD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UYVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;
 XX
 DR WPI: 2000-672631/65.
 DR P-PSDB; AAB36349.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 XX which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TFAP) nucleation depended
 CC assembly system of strains of Salmonella, Escherichia coli and
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Bacteriophage* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrillar protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention.

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.3e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
QY 61 GCGGTGTTCCACAATGGGGGGGGGGGTAATATACCGCGCGCAATAGTTCCGCG 120
DB 61 GCGGTGTTCCACAATGGGGGGGGGGGTAATATACCGCGCGCAATAGTTCCGCG 120
QY 121 CCGGACTATGATGACCTGTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
DB 121 CCGGACTATGATGACCTGTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
QY 181 AGCGATGCCGTAATCTGAACGACATTCACCAAGCGGTTAAGTAAACGCGCGCAT 240
DB 181 AGCGATGCCGTAATCTGAACGACATTCACCAAGCGGTTAAGTAAACGCGCGCAT 240
QY 241 GTAGGCGAGGTCGCGTAATATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGCGTAATATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGACGTAACGCTAAACCTCCGATTTACTGTCGCAATACGCGGT 360
DB 301 GCCACCATGACGACGTAACGCTAAACCTCCGATTTACTGTCGCAATACGCGGT 360
QY 361 AATAACGCGCGGTGTTTATGTAACGCGATCTGTAATGTAATGTAATGTAATGTAAT 420
DB 361 AATAACGCGCGGTGTTTATGTAACGCGATCTGTAATGTAATGTAATGTAATGTAAT 420
QY 421 GGTGTTGGCAACAGCGCGCTTAACGATTTAA 456
DB 421 GGTGTTGGCAACAGCGCGCTTAACGATTTAA 456

```

RESULT 2

AA087467 standard; DNA; 456 BP.

AA087467;
25-MAR-2003 (revised)
26-JUN-1995 (first entry)
Agfa sequence.
Salmonella; Agfa; vaccine; genetic immunization; ds.
Salmonella.

Key Location/Qualifiers
CDS 1..454
/tag=a
/note="Agfa"

W09425598-A2.

10-NOV-1994.

26-APR-1994; 94MO-IB000207.

26-APR-1993; 93US-00054452.

(U)VI-1 UNIV VICTORIA INNOVATION & DEV CORP.
(KING/) KING J.

Kay MW, Collinson SK, Clouthier SC, Doran JL,
WPI, 1994-358275/44.
P-PSDB; AAR74625.

Eliciting an immune response to *Salmonella* - using attenuated *Salmonella* strains, vector constructs, or compans. contg. fibrillar type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated proteins are used in genetic immunization and vaccine compositions, respectively, to elicit an immune response to *Salmonella* in animals (e.g. food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 2; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTCGACCAATCGTAGTTCGTGCGAGTCTGGCT 60
QY 61 GCGGTGTTCCACAATGGGGGGGGGGGTAATATACCGCGCGCAATAGTTCCGCG 120
DB 61 GCGGTGTTCCACAATGGGGGGGGGGGTAATATACCGCGCGCAATAGTTCCGCG 120
QY 121 CCGGACTATGATGACCTGTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
DB 121 CCGGACTATGATGACCTGTGTTACCGGTGTTTACCAGTAATAGSCAATGACATGCA 180
QY 181 AGCGATGCCGTAATCTGAACGACATTCACCAAGCGGTTAAGTAAACGCGCGCAT 240
DB 181 AGCGATGCCGTAATCTGAACGACATTCACCAAGCGGTTAAGTAAACGCGCGCAT 240
QY 241 GTAGGCGAGGTCGCGTAATATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGCGTAATATGTAATCTGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGACGTAACGCTAAACCTCCGATTTACTGTCGCAATACGCGGT 360
DB 301 GCCACCATGACGACGTAACGCTAAACCTCCGATTTACTGTCGCAATACGCGGT 360
QY 361 AATAACGCGCGGTGTTTATGTAACGCGATCTGTAATGTAATGTAATGTAATGTAAT 420
DB 361 AATAACGCGCGGTGTTTATGTAACGCGATCTGTAATGTAATGTAATGTAATGTAAT 420
QY 421 GGTGTTGGCAACAGCGCGCTTAACGATTTAA 456
DB 421 GGTGTTGGCAACAGCGCGCTTAACGATTTAA 456

```

RESULT 3

AAT74142

```

ID  AAT74142 standard; DNA; 456 BP.
XX
XX  AAT74142;
AC
XX  25-MAR-2003 (revised)
DT  29-SEP-1997 (first entry)
XX
DE  Salmonella enteritidis 27655-3b agfa gene.
XX
XX  Enteropathogenic bacteria; enterobacteria; S. enteritidis; antibody; ds.
XX
XX  Salmonella enteritidis.
OS
XX  Key      Location/Qualifiers
XX  CDS      1..456
XX            /tag= a
XX            /label= agfa_gene_fragment
XX            /transl_except= (pos:367..369,aa:Pro)
XX
XX  US5635617-A.
XX
XX  03-JUN-1997.
XX
XX  26-APR-1994; 94US-00233788.
XX
XX  26-APR-1993; 93US-00054452.
XX
XX  (UYVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX
XX  Collinson SK, Kay WW, Doran JL;
XX
XX  MPI: 1997-309886/28.
XX  P-PSDB; AAM23570.
XX
XX  Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
XX  enteropathogenic bacteria of the Enterobacteria family.
XX
XX  Claim 1; Col 19-112; 85pp; English.
XX
XX  The present sequence represents an isolated agfa gene derived from
XX  Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
XX  diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
XX  family Enterobacteria. It can also be used to provide proteins and
XX  antibodies which can be used for assays. The nucleic acid sequence can be
XX  used to provide probes or primers which can specifically hybridise to
XX  nucleic acid molecules from greater than 99% of Salmonella strains that
XX  are pathogenic to warm-blooded animals relative to nucleic acid molecules
XX  from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
XX  correct PF field.)
XX
XX  Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other:
SQ
Query Match      89.1%; Score 406.4; DB 2; Length 456;
Best Local Similarity 93.2%; Pred. No. 9,9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

```

DB  241 GTAGCCAGAGGTCGGATTAATAGTACTATTGAACGACTCAGAAATGTTTCGAAATAT 300
    |||
QY  301 GCCACCATCGACCACTGGAAACGCTAAAACTCCGATATTACTGTGCGCAATACGCGGT 360
    |||
DB  301 GCCACCATCGACCACTGGAAACGCTAAAACTCCGATATTACTGTGCGCAATACGCGGT 360
    |||
QY  361 AATAACGCGCGCTGTTTATATGACCCGATTTGATTCGAGGTAAAGTGTGCTCAGTT 420
    |||
DB  361 AATAACGCGCGCTGTTTATATGACCCGATTTGATTCGAGGTAAAGTGTGCTCAGTT 420
    |||
QY  421 GGTTTGGCAACACGCGCACTAACCAAGTATTA 456
    |||
DB  421 GGTTTGGCAACACGCGCACTAACCAAGTATTA 456
    |||

RESULT 4
AAC64617
ID  AAC64617 standard; DNA; 456 BP.
XX
XX  AAC64617;
XX
XX  26-FEB-2001 (first entry)
XX
XX  Salmonella enteritidis Agfa DNA sequence SEQ ID NO:1.
XX
XX  Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;
XX  vaccine; immune response; immunogen; ds.
XX
XX  Salmonella enteritidis.
XX
XX  WO200060102-A2.
XX
XX  12-OCT-2000.
XX
XX  05-APR-2000; 2000WO-CA000356.
XX
XX  05-APR-1999; 99US-0127888P.
XX
XX  (UYVI-) UNIV VICTORIA.
XX
XX  White AP, Doran JL, Collinson SK, Kay WW;
XX
XX  MPI: 2000-672631/65.
XX  P-PSDB; AAB36341.
XX
XX  Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX  which encodes foreign epitope or antigen, expresses recombinant Agfa
XX  protein useful for eliciting immune response in animal.
XX
XX  Disclosure: Page 134; 139pp; English.
XX
XX  The present invention describes a recombinant agfa gene (I) where a
XX  segment of the gene has been replaced by a segment of a foreign DNA
XX  sequence which encodes a foreign epitope or antigen. Also described are:
XX  (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended
XX  assembly system of strains of Salmonella, Bacterichia coli and
XX  Enterobacteriaceae for the production of fimbriae comprising recombinant
XX  Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX  the directing recombination of a recombinant gene into the chromosome of the
XX  homologous species; (3) directing recombination of a recombinant gene
XX  back into the chromosome of the homologous species, replacing the native
XX  copy of that gene; and (4) eliciting an immune response in an animal,
XX  comprising separating an amino acid polymer comprising a recombinant Agfa
XX  protein containing a replacement segment or segments of foreign amino
XX  acid sequence or sequences grown on a Salmonella, E. coli or
XX  Enterobacteriaceae host cell, from the host cell and introducing the
XX  polymer into the animal in conjunction with a carrier or diluent. (I) is
XX  useful for the expression of recombinant Agfa protein which is useful for
XX  eliciting an immune response in an animal. In a fimbrial presentation
XX  system the heterologous antigens are presented in high numbers (up to
XX  500,000 copies/cell), the hybrid fimbria protein possesses both the
XX  immunogenicity and adhesion properties relevant for an efficient live

```

CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
XX the exemplification of the present invention

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 89.1%; Score 406.4; DB 3; Length 456;
Best Local Similarity 93.2%; Pred. No. 9.9e-120;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGATTCGAGCAATGTAATTTCTGCGAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTCGAGCAATGTAATTTCTGCGAGTCTCTGCT 60
QY 61 GCGGCTGTTCCACATGGGCGGCGGATTAATCATAGCGGCGGCAATAGTTCGCG 120
DB 61 GCGGCTGTTCCACATGGGCGGCGGATTAATCATAGCGGCGGCAATAGTTCGCG 120
QY 121 CCGGACTATGATGAGTGTGTTACCCGTTGTTTACCCATGAATGCAATGCACTGCAA 180
DB 121 CCGGACTATGATGAGTGTGTTACCCGTTGTTTACCCATGAATGCAATGCACTGCAA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGCGGTTATGTTAGCGGCGGAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGCGGTTATGTTAGCGGCGGAT 240
QY 241 GTAGGCGGAGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGGAGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGACGATGGAAGCTTAAATCTGCAATTAATGTCGCGCAATGCGCGGT 360
DB 301 GCCACCATGACGACGATGGAAGCTTAAATCTGCAATTAATGTCGCGCAATGCGCGGT 360
QY 361 AATAACGCGCGGCTGTTAATCAGACCGCATGATTCGCGCAATGCGCGGT 420
DB 361 AATAACGCGCGGCTGTTAATCAGACCGCATGATTCGCGCAATGCGCGGT 420
QY 421 GGTGTTGGCAACAGCGCAGCGCTTAACCAAGTATTA 456
DB 421 GGTGTTGGCAACAGCGCAGCGCTTAACCAAGTATTA 456

```

RESULT 5

AA64626 standard; DNA; 456 BP.

AA64626;

26-FEB-2001 (first entry)

Agfa::PT3#5 DNA sequence SEQ ID NO:19.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope;

Escherichia coli.

Synthetic.

MO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000WO-CA000356.

05-APR-1999; 99US-0127888P.

(UVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay RW;

WP: 2000-672631/65.
P-PSDB: AAB36350.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure, Page 137, 139pp; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA (1) use of thin aggregative fimbriae (SEF17/TAH) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant directing recombinant of a recombinant subunits, respectively; (2) Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (3) homologous species; (3) directing recombinant gene into the chromosome of the back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant with a carrier which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 79.6%; Score 363.2; DB 3; Length 456;
Best Local Similarity 87.3%; Pred. No. 6.9e-106;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGCAGATTCGAGCAATGTAATTTCTGCGAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTCGAGCAATGTAATTTCTGCGAGTCTCTGCT 60
QY 61 GCGGCTGTTCCACATGGGCGGCGGATTAATCATAGCGGCGGCAATAGTTCGCG 120
DB 61 GCGGCTGTTCCACATGGGCGGCGGATTAATCATAGCGGCGGCAATAGTTCGCG 120
QY 121 CCGGACTATGATGAGTGTGTTACCCGTTGTTTACCCATGAATGCAATGCACTGCAA 180
DB 121 CCGGACTATGATGAGTGTGTTACCCGTTGTTTACCCATGAATGCAATGCACTGCAA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGCGGTTATGTTAGCGGCGGAT 240
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTACCCAGCGGTTATGTTAGCGGCGGAT 240
QY 241 GTAGGCGGAGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGGAGGCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACCATGACGACGATGGAAGCTTAAATCTGCAATTAATGTCGCGCAATGCGCGGT 360
DB 301 GCCACCATGACGACGATGGAAGCTTAAATCTGCAATTAATGTCGCGCAATGCGCGGT 360
QY 361 AATAACGCGCGGCTGTTAATCAGACCGCATGATTCGCGCAATGCGCGGT 420
DB 361 AATAACGCGCGGCTGTTAATCAGACCGCATGATTCGCGCAATGCGCGGT 420
QY 421 GGTGTTGGCAACAGCGCAGCGCTTAACCAAGTATTA 456
DB 421 GGTGTTGGCAACAGCGCAGCGCTTAACCAAGTATTA 456

```


DB 421 GGTITGGCAACACGCCAGCTAACGATATTAA 456

RESULT 6
AAC64628
ID AAC64628 standard; DNA; 456 BP.
AC AAC64628;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX
XX Salmone11a: agfa: chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
PD 05-APR-2000; 2000WO-CA000356.
PF 05-APR-1999; 99US-0127888P.
XX 05-APR-1999; 99US-0127888P.
XX (UTVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
DR P-PSDB; AAB36352.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure: Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/TRF) nucleation depended assembly system of strains of Salmone11a, Escherichia coli and Enterobacteriaceae for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2) the directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmone11a, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbriin protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.5e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCAATTCGACGAATCGTAGTTTCGAGTGCCTGGCT 60
DB 1 ATGAAACTTTTAAAGTGGCAGCAATTCGACGAATCGTAGTTTCGAGTGCCTGGCT 60
QY GGCCTGTTCCACAATGGGGCGGCGCGGTATCATACGCGCGGCAATAGTCCGGC 120
DB GGCCTGTTCCACAATGGGGCGGCGCGGTATCATACGCGCGGCAATAGTCCGGC 120
QY 121 CCGGACTATGATCAGCTGCTTACCCGTTGTTCACCATGAATGCGCATCTGCAA 180
DB 121 CCGGACTATGATGAGCATTTATCAGTACCGTTCCGTTACCGCTGCTCTGCAA 180
QY 181 AGCGATGCCGCTTAATCGAATGCAATTCACGACGCGGTTATGTAACGCGCGCAT 240
DB 181 AGCGATGCCGCTTAATCGAATGCAATTCACGACGCGGTTATGTAACGCGCGCAT 240
QY 241 GTAGGCGAGGTGGGATATAGTACTTATGAACTGACTCAGAAATGTTTCAAAATAT 300
DB 241 GTAGGCGAGGTGGGATATAGTACTTATGAACTGACTCAGAAATGTTTCAAAATAT 300
QY 301 GCCACCATCGACCAATGGAACGCTTAAATCTCGATATTACTGCGCAATACGCGGT 360
DB 301 GCCACCATCGACCAATGGAACGCTTAAATCTCGATATTACTGCGCGGTACCGGT 360
QY 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGTT 420
DB 361 CATGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATG 420
QY 421 GGTITGGCAACACGCCAGCTAACGATATTAA 456
DB 421 GGTITGGCAACACGCCAGCTAACGATATTAA 456

RESULT 7
AAC64622
ID AAC64622 standard; DNA; 456 BP.
AC AAC64622;
XX
XX 26-FEB-2001 (first entry)
DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
XX
XX Salmone11a: agfa: chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO200060102-A2.
XX 12-OCT-2000.
PD 05-APR-2000; 2000WO-CA000356.
PF 05-APR-1999; 99US-0127888P.
XX 05-APR-1999; 99US-0127888P.
XX (UTVI-) UNIV VICTORIA.
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
DR P-PSDB; AAB36346.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure: Page 135; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC Enterobacteriaceae for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a Salmonella, E. coli or
 CC Enterobacteriaceae host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX

Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGATGTTCTGGCAGTGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGATGTTCTGGCAGTGTCTGGCT 60
 QY 61 GCGGCTGTCGCAAAATGGGCGGCGGCGGATATCAACGCGGCGGCAATGTTCCGGC 120
 DB 61 GCGGCTGTCGCAAAATGGGCGGCGGCGGATATCAACGCGGCGGCAATGTTCCGGC 120
 QY 121 CCGGACTATGATCAGCTGTTACCCGTTGTTACCCATGAATGCAATGCACTGCA 180
 DB 121 CCGGACTATGATCAGCTGTTACCCGTTGTTACCCATGAATGCAATGCACTGCA 180
 QY 181 AGCATGCGCGTAAATTTGAAAGCAATTCACAGCGGTTATGTAAGCGGCGCAT 240
 DB 181 AGCATGCGCGTAAATTTGAAAGCAATTCACAGCGGTTATGTAAGCGGCGCAT 240
 QY 241 GTAGGCGGAGGTGCGGATATGTAATTTGAAGTCAAGATGTTTCAAGAAATTA 300
 DB 241 GTAGGCGGAGGTGCGGATATGTAATTTGAAGTCAAGATGTTTCAAGAAATTA 300
 QY 301 GCCACCATGACGACATGGAAGCTTAAAGTCCGATATTAATCTGCGGCAATGCGCGT 360
 DB 301 GCCACCATGACGACATGGAAGCTTAAAGTCCGATATTAATCTGCGGCAATGCGCGT 360
 QY 361 AATTAAGCGCGGCTGTTATCAACGCAATGATTCGAGCTAATGGTGGTCAAGTT 420
 DB 361 AATTAAGCGCGGCTGTTATCAACGCAATGATTCGAGCTAATGGTGGTCAAGTT 420
 QY 421 GATTGGCAACAAGCGGCGGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 456
 DB 421 GCAATGCAACAAGCGGCGGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 456

RESULT 8

ID AAC64629 standard; DNA; 456 BP.

AC AAC64629;

XX 26-FEB-2001 (first entry)

DE Agfa::PT3#8 DNA sequence SEQ ID NO:25.

XX Salmonella; agfa; chromosome1 gene replacement; fimbrial; epitope;

XX vaccine; immune response; immunogen; ds.

XX Salmonella enteritidis.

XX Escherichia coli.

XX Synthetic.

XX WO200060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UUVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay W;

XX WPI: 2000-672631/65.

XX P-PSDB; AAB36353.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence

XX which encodes foreign epitope or antigen, expresses recombinant Agfa

XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 138; 139p; English.

XX The present invention describes a recombinant agfa gene (1) where a

XX segment of the gene has been replaced by a segment of a foreign DNA

XX sequence which encodes a foreign epitope or antigen. Also described are:

XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended

XX assembly system of strains of Salmonella, Escherichia coli and

XX Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)

XX directing recombination of a recombinant gene into the chromosome of the

XX homologous species; (3) directing recombination of a recombinant gene

XX back into the chromosome of the homologous species, replacing the native

XX copy of that gene; and (4) eliciting an immune response in an animal,

XX protein containing a replacement segment or segments of foreign amino

XX acid sequence or sequences grown on a Salmonella, E. coli or

XX Enterobacteriaceae host cell, from the host cell and introducing the

XX polymer into the animal in conjunction with a carrier or diluent. (1) is

XX useful for the expression of recombinant Agfa protein which is useful for

XX eliciting an immune response in an animal. In a fimbrial presentation

XX system the heterologous antigens are presented in high numbers (up to

XX 500,000 copies/cell), the hybrid fimbrial protein possesses both the

XX immunogenicity and adhesion properties relevant for an efficient live

XX vaccine, the carrier fimbrial subunit proteins are usually strong

XX immunogens, which may be important for directing an immune response

XX against the inserted epitope, and hybrid fimbriae are easy and

XX inexpensive to purify in large amount. The present sequence is given in

XX the exemplification of the present invention

QY Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match Best Local Similarity 77.9%; Score 355.2; DB 3; Length 456;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGATGTTCTGGCAGTGTCTGGCT 60
 DB 1 ATGAACTTTTAAAGTGCGAGCATTCGACGATTCGATGTTCTGGCAGTGTCTGGCT 60
 QY 61 GCGGCTGTCGCAAAATGGGCGGCGGCGGATATCAACGCGGCGGCAATGTTCCGGC 120
 DB 61 GCGGCTGTCGCAAAATGGGCGGCGGCGGATATCAACGCGGCGGCAATGTTCCGGC 120
 QY 121 CCGGACTATGATCAGCTGTTACCCGTTGTTACCCATGAATGCAATGCAATGCAAT 180
 DB 121 CCGGACTATGATCAGCTGTTACCCGTTGTTACCCATGAATGCAATGCAATGCAAT 180

```
DB 121 CCGAGCTCAACGTTGAGCATTTATCATGACGTTCCGCTAACGTCGCTTATGATCAG 180
QY 181 AGCGATGCCCGCTTAATCTGAAACGACATTTACCCAGAGGGTTATGTTACGGCCCGCAT 240
DB 181 CTGGTTACCCCGTGTGTTATCCCATGAAATGAGCATGACGGTTATGTTACGGCCCGCAT 240
QY 241 GTAGGCCAGGGTGGCGATATAGTACTATTTGAACTGACACTCAGATGTTGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATAGTACTATTTGAACTGACACTCAGATGTTGTTTCAAAATAT 300
QY 301 GCCACCATGACACGACGTGAACCGCTAAACCTCCGATATTTACTGTCGCCCAATACGGCGGT 360
DB 301 GCCACCATGACACGACGTGAACCGCTAAACCTCCGATATTTACTGTCGCCCAATACGGCGGT 360
QY 361 AATAACGGCGGCTGTTTATATGACACCGATCTGATTTCCAGGTTATGTTGCTCAGGTT 420
DB 361 AATAACGGCGGCTGTTTATATGACACCGATCTGATTTCCAGGTTATGTTGCTCAGGTT 420
QY 421 GGTTCGCAACCAACGACGCGCTAACGATTTAA 456
DB 421 GGTTCGCAACCAACGACGCGCTAACGATTTAA 456

RESULT 9
AAC64623
ID AAC64623 standard; DNA; 456 BP.
AC AAC64623;
DT 26-FEB-2001 (first entry)
DE Agfa::PT3#2 DNA sequence SEQ ID NO:13.
KM Salmone11a; agfa; chromosomal gene replacement; fimbria; epitope;
OS vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO20060102-A2.
XX 12-OCT-2000.
XX 05-APR-2000; 2000WO-CA000356.
XX 05-APR-1999; 99US-0127888P.
XX (UYVI-) UNIV VICTORIA.
XX White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.
XX P-PSDB; AAB36347.
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
XX which encodes foreign epitope or antigen, expresses recombinant Agfa
XX protein useful for eliciting immune response in animal.
XX Disclousure; Page 136; 139p; English.
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described
XX (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
XX assembly system of strains of Salmone11a, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
```

```
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmone11a, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (1) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention
SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;
Query Match 77.9%; Score 355.2; DB 3; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.5e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
QY 1 ATGAAACTTTTAAAGTGGACGATTCGACCAATCGATGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAAACTTTTAAAGTGGACGATTCGACCAATCGATGTTCTGGCAGTCTCTGCT 60
QY 61 GCGCTCTTCCCAATAGGGGGCGGCGGCTATATCAATACGGCGGCGCAATAGTCCGCG 120
DB 61 GCGCTCTTCCCAATAGGGGGCGGCGGCTATATCAATACGGCGGCGCAATAGTCCGCG 120
QY 121 CCGACTATGATCAGCTGTTTACCGGTGTGTTTACCCATGAAATGACATGACATGCA 180
DB 121 CCGACTATGATGAGATTTATCATGACGTTCCGTTACGCTGCTGCTGCTGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTTACCAAGACGCGTTATGTTACGGCGCGAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTTACCAAGACGCGTTATGTTACGGCGCGAT 240
QY 241 GTAGGCCAGGGTGGCGATATAGTACTATTTGAACTGACACTCAGATGTTTCAAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATATAGTACTATTTGAACTGACACTCAGATGTTTCAAAATAT 300
QY 301 GCCACCATGACACGACGTGAACCGCTAAACCTCCGATATTTACTGTCGCCCAATACGGCGGT 360
DB 301 GCCACCATGACACGACGTGAACCGCTAAACCTCCGATATTTACTGTCGCCCAATACGAC 360
QY 361 AATAACGGCGGCTGTTTATATGACACCGATCTGATTTCCAGGTTATGTTGCTCAGGTT 420
DB 361 CTGGTTACCCCGTGTGTTATCCCATGAAATGAGCATGACGTTATGTTGCTCAGGTT 420
QY 421 GGTTCGCAACCAACGACGCGCTAACGATTTAA 456
DB 421 GGTTCGCAACCAACGACGCGCTAACGATTTAA 456

RESULT 10
AAC64624
ID AAC64624 standard; DNA; 456 BP.
AC AAC64624;
DT 26-FEB-2001 (first entry)
DE Agfa::PT3#3 DNA sequence SEQ ID NO:15.
KM Salmone11a; agfa; chromosomal gene replacement; fimbria; epitope;
OS vaccine; immune response; immunogen; ds.
OS Salmone11a enteritidis.
OS Escherichia coli.
OS Synthetic.
XX WO20060102-A2.
XX
```


CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2.7e-102;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGCGACGATTCGAGCAATCGATGTTCTGCGAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGCGACGATTCGAGCAATCGATGTTCTGCGAGTCTGCT 60
QY 61 GCGCTGTTTCCACATGGGCGCGCGGTTATCTAAACGGCGCGCAATGTTCCGCG 120
DB 61 GCGCTGTTTCCACATGGGCGCGCGGTTATCTAAACGGCGCGCAATGTTCCGCG 120
QY 121 CCGGACTATGATCACTGGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCGAA 180
DB 121 CCGGACTATGATCACTGGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCGAA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCGCGAT 240
QY 241 GTAGCCGAGGTCGCGATTAATGTAATCTTGAATGATCTGATGATGTTTCAAAATAT 300
DB 241 GTAGCCGAGGTCGCGATTAATGTAATCTTGAATGATCTGATGATGTTTCAAAATAT 300
QY 301 GCCACATCGACGACGATGGAACGCTTAAATCCGATTTATCTGCGCGCAATAGCGCGGT 360
DB 301 GCCACATCGACGACGATGGAACGCTTAAATCCGATTTATCTGCGCGCAATAGCGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTTGCGTCAAGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTTGCGTCAAGTT 420
QY 421 GGTITGGCAACAAAGCGGCTTAACCAATTTAA 456
DB 421 GGTITGGCAACAAAGCGGCTTAACCAATTTAA 456

```

RESULT 12

AAC64627 standard; DNA; 456 BP.

XX AAC64627;

XX 26-FEB-2001 (first entry)

XX Agfa::PT3#6 DNA sequence SEQ ID NO:21.

XX Salmomella; agfa; chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen; ds.

XX Salmomella enteritidis.

XX Escherichia coli.

XX Synthetic.

XX WO20060102-A2.

XX 12-OCT-2000.

XX 05-APR-2000; 2000WO-CA000356.

XX 05-APR-1999; 99US-0127888P.

XX (UVI-) UNIV VICTORIA.

XX White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SFP1/TFAP) nucleation depended
 CC assembly system of strains of *Salmomella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmomella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (I) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 77.2%; Score 352; DB 3; Length 456;
 Best Local Similarity 85.7%; Pred. No. 2.7e-102;
 Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGCGACGATTCGAGCAATCGATGTTCTGCGAGTCTGCT 60
DB 1 ATGAACTTTTAAAGTGCGACGATTCGAGCAATCGATGTTCTGCGAGTCTGCT 60
QY 61 GCGCTGTTTCCACATGGGCGCGCGGTTATCTAAACGGCGCGCAATGTTCCGCG 120
DB 61 GCGCTGTTTCCACATGGGCGCGCGGTTATCTAAACGGCGCGCAATGTTCCGCG 120
QY 121 CCGGACTATGATCACTGGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCGAA 180
DB 121 CCGGACTATGATCACTGGTTACCGGTGTTTACCCATGAAATGGCAATGCACTGCGAA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAAACGACATTCACGAGCGGTTATGTTAAGCGCGCGAT 240
QY 241 GTAGCCGAGGTCGCGATTAATGTAATCTTGAATGATCTGATGATGTTTCAAAATAT 300
DB 241 GTAGCCGAGGTCGCGATTAATGTAATCTTGAATGATCTGATGATGTTTCAAAATAT 300
QY 301 GCCACATCGACGACGATGGAACGCTTAAATCCGATTTATCTGCGCGCAATAGCGCGGT 360
DB 301 GCCACATCGACGACGATGGAACGCTTAAATCCGATTTATCTGCGCGCAATAGCGCGGT 360
QY 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTTGCGTCAAGTT 420
DB 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTTGCGTCAAGTT 420
QY 421 GGTITGGCAACAAAGCGGCTTAACCAATTTAA 456
DB 421 GGTITGGCAACAAAGCGGCTTAACCAATTTAA 456

```

1 ATGAACCTTTTAAAGTGCAGCATTCGACGAATCGTAGTTCTTGGCAGTGCCTCTGGCT 60

26-APR-1994; 94WO-IB00020

Thu Mar 18 12:28:00 2004

Job time : 246.684 secs

us-09-543-407-17.rng

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds

(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttcaaaagtcgc.....ccagcgtcaaccagatratraa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents NA:*

```
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCCTS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406.4	89.1	456	1	US-08-233-788A-58
2	282.2	61.9	361	1	US-08-233-788A-56
3	36.4	8.0	798	4	US-09-252-991A-13717
4	36.4	8.0	2748	4	US-09-252-991A-13830
5	36.4	8.0	3720	4	US-09-252-991A-13627
6	33.2	7.3	4403765	3	US-09-103-840A-2
7	33.2	7.3	4411529	3	US-09-103-840A-1
8	33.2	7.2	834	4	US-09-621-976-2574
9	32.4	7.1	332	3	US-08-469-412A-8
10	32.4	7.1	332	3	US-09-021-715-8
11	31.4	6.9	2007	4	US-09-643-597-153
12	31.4	6.9	2007	4	US-09-480-884A-153
13	31.4	6.9	2007	4	US-09-542-615A-153
14	31.4	6.9	2007	4	US-09-606-421B-153
15	31.4	6.9	2007	4	US-09-221-107-153
16	31.4	6.9	2148	4	US-09-643-597-154
17	31.4	6.9	2148	4	US-09-480-884A-154
18	31.4	6.9	2148	4	US-09-542-615A-154
19	31.4	6.9	2148	4	US-09-606-421B-154
20	31.4	6.9	2148	4	US-09-221-107-154
21	31.2	6.8	1212	4	US-09-252-991A-7452
22	31.2	6.8	1425	4	US-09-252-991A-7188
23	31.2	6.8	1491	6	5466473-3
24	31.2	6.8	1692	4	US-09-252-991A-7513
25	31.2	6.8	2436	4	US-09-540-236-492
26	31.2	6.8	65792	4	US-09-596-002-31
27	30.8	6.8	2547	3	US-08-508-761B-1

28	30.8	6.8	1664976	4	US-08-916-421B-1	Sequence 1, Appl
29	30.4	6.7	552	4	US-09-252-991A-9800	Sequence 9800, Ap
30	30.4	6.7	894	4	US-09-434-354-1	Sequence 1, Appl
31	30.2	6.6	774	4	US-09-252-991A-11893	Sequence 11893, A
32	30.2	6.6	1083	4	US-09-252-991A-11930	Sequence 11930, A
33	30.2	6.6	3323	2	US-08-422-699A-10	Sequence 10, Appl
34	30.2	6.6	3323	2	US-08-422-706B-10	Sequence 10, Appl
35	29.8	6.5	1500	4	US-09-252-991A-6254	Sequence 6254, Ap
36	29.8	6.5	1695	4	US-09-252-991A-6044	Sequence 6044, Ap
37	29.8	6.5	1747	4	US-09-566-921-56	Sequence 66, Appl
38	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
39	29.4	6.4	832	4	US-09-621-976-2813	Sequence 2813, Ap
40	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
41	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
42	29.4	6.4	1252	4	US-09-954-687-29	Sequence 29, Appl
43	29.4	6.4	1939	4	US-08-961-527-310	Sequence 310, Ap
44	29.4	6.4	3494	3	US-09-139-802-200	Sequence 200, Ap
45	29.4	6.4	3494	4	US-09-659-786-200	Sequence 200, Ap

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 3633617
GENERAL INFORMATION:
APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Sharon S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELETYPE: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58

Query Match 89.1%; Score 406.4; DB 1; Length 456;
Best Local Similarity 93.2%; Pred. No. 3.1e-124;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGAATCGTAGTTCTGGAGTCTCGGCT 60
 Db 1 ATGAACTTTTAAAGTGGACGATTCGACGAATCGTAGTTCTGGAGTCTCGGCT 60
 QY 61 GGGCTGCTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
 Db 61 GGGCTGCTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
 QY 121 CCGGACTATGATCAGCTGGTTACCCGTGTGTTGTTACCCATGAATGCGACATGACGCA 180
 Db 121 CCGGACTATGATCAGCTGGTTATCAGTAGCGTTCCGCTAACGGTGGCTTCTGCA 180
 QY 181 AGCGATGCGCGTAAATCGGAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 240
 Db 181 AGCGATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 240
 QY 241 GTAGGCGGAGGTGGGATTAATAGTACTATTAAGTGAAGTGAAGTGAAGTGAAGTGA 300
 Db 241 GTAGGCGGAGGTGGGATTAATAGTACTATTAAGTGAAGTGAAGTGAAGTGAAGTGA 300
 QY 301 GCCACCATTCACCGAGTGAACGCTTAACTCCGATTTAATCTGTCGCGCAATACGGCGGT 360
 Db 301 GCCACCATTCACCGAGTGAACGCTTAACTCCGATTTAATCTGTCGCGCAATACGGCGGT 360
 QY 361 AATAACGCGCGCTGTTATATCAGACGCGATCTGATTCAGAGGTTATGTTGCGTCAAGTT 420
 Db 361 AATAACGCGCGCTGTTATATCAGACGCGATCTGATTCAGAGGTTATGTTGCGTCAAGTT 420
 QY 421 GGTGTTGGCAACAACCGCGCTAACCGATTAACCGATTAACCGATTAACCGATTAAC 456
 Db 421 GGTGTTGGCAACAACCGCGCTAACCGATTAACCGATTAACCGATTAACCGATTAAC 456

RESULT 2
 US-08-233-788A-56
 Sequence 56, Application US/08233788A
 Patent No. 5635617
 GENERAL INFORMATION:
 APPLICANT: Doran, James L.
 APPLICANT: Kay, William W.
 APPLICANT: Clouthier, Karen S.
 APPLICANT: Clouthier, Sharon C.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
 NUMBER OF SEQUENCES: 61
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Seed and Berry
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: U.S.A.
 ZIP: 98104-7092
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/233,788A
 FILING DATE: 26-APR-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: King, Joshua
 REGISTRATION NUMBER: 35,570
 REFERENCE/DOCKET NUMBER: 920043,403C2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031
 TELEX: 3723836 SEEDANBERRY
 INFORMATION FOR SEQ ID NO: 56:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 361 base pairs

TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 1..357
 US-08-233-788A-56

Query Match 61.9%; Score 282.2; DB 1; Length 361;
 Best Local Similarity 90.1%; Pred. No. 2,5e-83;
 Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 64 GTGCTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123
 Db 1 GTGCTTCCCAATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 123
 QY 124 GACTATGATCAGCTGGTTACCCGTGTGTTGTTACCCATGAATGCGACATGCAAGC 183
 Db 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACCGCTGCTCTGCAAAAGC 120
 QY 184 GATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCATGTA 243
 Db 121 GATGCGCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCATGTA 180
 QY 244 GGGCAGGCTGGGATTAATAGTACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 303
 Db 181 GGGCAGGCTGGGATTAATAGTACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 240
 QY 304 ACCATGACCATGGAACGCTTAACTCCGATTTAATCTGTCGCGCAATACGGCGGTAT 363
 Db 241 ACCATGACCATGGAACGCTTAACTCCGATTTAATCTGTCGCGCAATACGGCGGTAT 300
 QY 364 AACGCGCGCTGTTATATCAGACGCGATCTGATTC 398
 Db 301 AACGCGCGCTGTTATATCAGACGCGATCTGATTC 335

RESULT 3
 US-09-252-991A-13717/c
 Sequence 13717, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 PRIOR FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13717
 LENGTH: 798
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13717

Query Match 8.0%; Score 36.4; DB 4; Length 798;
 Best Local Similarity 47.4%; Pred. No. 0.04;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGATTCGACGAATCGTAGTTCTGGAGTCTCGGCTGTTCCACAT 76
 Db 270 TTGGCGGCTCGGCGAGCTGTGAAGGCGTGAACATCGACGAGGTGCTTCACTGT 211
 QY 77 GGGGCGGCGCGGTTATCAATACGCGCGGCAATAGTTCGCGCGCATATGATCAGC 136
 Db 210 TCGGCGGCGGTTGTGCGAGTTCGCGGTGTCTCAGACAGTGCACAGGATGCGATGTTG 151
 QY 137 TGGTACCCGCTGTGTTACCCATGAATGGAATGACATGACATGCAAGCGATGCCGTAAAT 196

Db 150 TAGCCGCCCTGATGGTGCACAGCTGGCGGACGTTGCTTGCATTAAGGGGGAG 91
 QY 197 CTGAACGACCATTAACCCAGACGGTTATGTTAGCGGCCCATATGAGC 246
 Db 90 GTGGCTTGCCTTGACGATGCGGAAGAAACGGCCTTGACGTAGGC 41

RESULT 4 US-09-252-991A-13820/C

Sequence 13820, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13820
 LENGTH: 2748
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13820

Query Match 8.0%; Score 36.4; DB 4; Length 2748;
 Best Local Similarity 47.4%; Pred. No. 0.074;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGCATTCGACGAAATCGTAGTTTCTGCGAGTCTCTGCGCGCTGCTTCCACAT 76
 Db 548 TTGGCGCGCTCGGACCGTCGTGAAGCGCGTGAACATGACGAGGTGTGCTTCACTGT 489
 QY 77 GGGGGGGGGGGGGTATATATACGGCGGCGCAATAGTTCCGGCCCGCATATATAC 136
 Db 488 TCGGGCGGGGTGTGGCCAGTTTGGCGGTCTGTCACAGCTCCACGAGGCGTGGCATGTTG 429
 QY 137 TGGTACCGGTGTGTATACCATGAAATGACATGACATGCAAGCCGATGCCGTAAT 196
 Db 428 TAGCCGCCCTGATGCTGCGCAGCACTGCGGGCAGCATGCTTGTTCATTAAGGGGGAG 369
 QY 197 CTGAACGACCATTAACCCAGACGGTTATGTTAGCGGCGCCGATGAGC 246
 Db 368 GTGGCTTGCCTTGACGATGCGGAAGAAACGGCCTTGACGTAGGC 319

RESULT 5

US-09-252-991A-13627
 Sequence 13627, Application US/09252991A
 Patent No. 6551795
 GENERAL INFORMATION:
 APPLICANT: Marc J. Rubenfield et al.
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 FILE REFERENCE: 107196.136
 CURRENT APPLICATION NUMBER: US/09/252,991A
 CURRENT FILING DATE: 1999-02-18
 PRIOR APPLICATION NUMBER: US 60/074,788
 PRIOR FILING DATE: 1998-02-18
 PRIOR APPLICATION NUMBER: US 60/094,190
 PRIOR FILING DATE: 1998-07-27
 NUMBER OF SEQ ID NOS: 33142
 SEQ ID NO 13627
 LENGTH: 3720
 TYPE: DNA
 ORGANISM: Pseudomonas aeruginosa
 US-09-252-991A-13627

Query Match 8.0%; Score 36.4; DB 4; Length 3720;
 Best Local Similarity 47.4%; Pred. No. 0.087;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 17 TGGCAGCATTCGACGAAATCGTAGTTTCTGCGAGTCTCTGCGCGCTGCTTCCACAT 76
 Db 3080 TTGGCGCGCTCGGAGACGTCGTGAAGCGCTGCAACATCAGAGGTGTGCTTCACTGT 3139
 QY 77 GGGGGGGGGGGGGTATATATACGGCGGCGCAATAGTTCCGGCCCGCATATATAC 136
 Db 3140 TCGGGCGGGGTGTGGCCAGTTTGGCGGTCTGTCACAGCTCCACGAGGCGGCGATGTTG 3199
 QY 137 TGGTACCGGTGTGTATACCATGAAATGACATGACATGCAAGCCGATGCCGTAAT 196
 Db 3200 TAGCCGCCCTGATGCTGCCAGCACTGCGCGGACGCTTGTCTGTCAATTAAGGGGGAG 3259
 QY 197 CTGAACGACCATTAACCCAGACGGTTATGTTAGCGGCGCCGATGAGC 246
 Db 3260 GTGGCTTGCCTTGACGATGCGGAAGAAACGGCCTTGACGTAGGC 3309

RESULT 6

US-09-103-840A-2
 Sequence 2, Application US/09103840A
 Patent No. 6294328
 GENERAL INFORMATION:
 APPLICANT: FLEISCHMAN, Robert D.
 APPLICANT: WHITE, Owen R.
 APPLICANT: FRASER, Claire M.
 APPLICANT: VENTER, John C.
 TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
 TITLE OF INVENTION: TUBERCULOSIS
 FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 CURRENT FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: Patent In Ver. 2.1
 SEQ ID NO 2
 LENGTH: 4403765
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: CDC 1551
 OTHER INFORMATION: "n" bases at various positions throughout the sequence
 OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Query Match 7.3%; Score 33.2; DB 3; Length 4403765;
 Best Local Similarity 44.6%; Pred. No. 31;
 Matches 131; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

QY 77 GGGGGGGGGGGGGTATATATACGGCGGCGCAATAGTTCCGGCCCGCATATATAC 136
 Db 1093443 GCGGTGCGGAGGTGATGCGGAGCCCGGTGACGCTGTGGTGTGCGGCTCATACCGAC 1093502
 QY 137 TGGTACCGGTGTGTATACCATGAAATGACATGACATGCAAGCGATCCGTAAT 196
 Db 1093503 ACCGTGTCTGTGAAATGCTACGAAAGCCGCGGTGACGAGGCTTACCGCAC 1093562
 QY 197 CTGAACGACCATTAACCCAGACGGTTATGTTAGCGGCGCCGATGAGCCAGGAGTGGC 256
 Db 1093563 CCGCTTGACCGCTTGCGGATGTATGCGGAGCGCGTGAAGCTGTGTGCGGCGGTTG 1093622
 QY 257 ATTAATGATTAATGAACTGACGTAATGTTTTCAGAAATTAATGCAACATGACCACT 316
 Db 1093623 ATCAACGACACCGTGGCGCTGCCAATATGCTGACGTAGCGGCTCGGCTTAC 1093682
 QY 317 GGAACGCTAAAACTCGGATTAATGCTGCGGCAATCGGCGGATTAATGAGCGC 370
 Db 1093683 GCCACCGCGTGGCGGATGTTGAAGCTGCGGCGGTATGTTGATGACGCGCCG 1093736

RESULT 7

Query Match	7.3%	Score 33.2;	DB 3;	Length 4411529;
Best Local Similarity	44.6%	Pred. No. 31;		
Matches 131;	Conservative	0;	Mismatches 173	

	1093444	1093445	1093446	1093447	1093448	1093449	1093450	1093451	1093452	1093453	1093454	1093455	1093456	1093457	1093458	1093459	1093460	1093461	1093462	1093463	1093464	1093465	1093466	1093467	1093468	1093469	1093470	1093471	1093472	1093473	1093474	1093475	1093476	1093477	1093478	1093479	1093480	1093481	1093482	1093483	1093484	1093485	1093486	1093487	1093488	1093489	1093490	1093491	1093492	1093493	1093494	1093495	1093496	1093497	1093498	1093499	1093500	1093501	1093502	1093503	1093504	1093505	1093506	1093507	1093508	1093509	1093510	1093511	1093512	1093513	1093514	1093515	1093516	1093517	1093518	1093519	1093520	1093521	1093522	1093523	1093524	1093525	1093526	1093527	1093528	1093529	1093530	1093531	1093532	1093533	1093534	1093535	1093536	1093537	1093538	1093539	1093540	1093541	1093542	1093543	1093544	1093545	1093546	1093547	1093548	1093549	1093550	1093551	1093552	1093553	1093554	1093555	1093556	1093557	1093558	1093559	1093560	1093561	1093562	1093563	1093564	1093565	1093566	1093567	1093568	1093569	1093570	1093571	1093572	1093573	1093574	1093575	1093576	1093577	1093578	1093579	1093580	1093581	1093582	1093583	1093584	1093585	1093586	1093587	1093588	1093589	1093590	1093591	1093592	1093593	1093594	1093595	1093596	1093597	1093598	1093599	1093600	1093601	1093602	1093603	1093604	1093605	1093606	1093607	1093608	1093609	1093610	1093611	1093612	1093613	1093614	1093615	1093616	1093617	1093618	1093619	1093620	1093621	1093622	1093623	1093624	1093625	1093626	1093627	1093628	1093629	1093630	1093631	1093632	1093633	1093634	1093635	1093636	1093637	1093638	1093639	1093640	1093641	1093642	1093643	1093644	1093645	1093646	1093647	1093648	1093649	1093650	1093651	1093652	1093653	1093654	1093655	1093656	1093657	1093658	1093659	1093660	1093661	1093662	1093663	1093664	1093665	1093666	1093667	1093668	1093669	1093670	1093671	1093672	1093673	1093674	1093675	1093676	1093677	1093678	1093679	1093680	1093681	1093682	1093683	1093684	1093685	1093686	1093687	1093688	1093689	1093690	1093691	1093692	1093693	1093694	1093695	1093696	1093697	1093698	1093699	1093700	1093701	1093702	1093703	1093704	1093705	1093706	1093707	1093708	1093709	1093710	1093711	1093712	1093713	1093714	1093715	1093716	1093717	1093718	1093719	1093720	1093721	1093722	1093723	1093724	1093725	1093726	1093727	1093728	1093729	1093730	1093731	1093732	1093733	1093734	1093735	1093736	1093737	1093738	1093739	1093740	1093741	1093742	1093743	1093744	1093745	1093746	1093747	1093748	1093749	1093750	1093751	1093752	1093753	1093754	1093755	1093756	1093757	1093758	1093759	1093760	1093761	1093762	1093763	1093764	1093765	1093766	1093767	1093768	1093769	1093770	1093771	1093772	1093773	1093774	1093775	1093776	1093777	1093778	1093779	1093780	1093781	1093782	1093783
--	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------	---------

```

RESULT 8
US-09-621-976-2574
; Sequence 2574 Application US/09621976 .
; Patent No 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J. B.
APPLICANT: Jobert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins
FILE REFERENCE: GENSET.054P2
CURRENT APPLICATION NUMBER: US/09/621,976
CURRENT FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent.pm
SEQ ID NO 2574
LENGTH: 834
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 497..826
US-09-621-976-2574

```

Query Match 7.2%; Score 33; DB 4; Length 834;
Best Local Similarity 14.9%; Pred. No. 0.54;
Matches 18; Conservative 64; Mismatches 39; Indels 0; Gaps 0

QY 79 GGGGGGGGGGTAATCATATACGGCGGGGCAATAGTCCGGCCCGCATATGATCAGCTG 13
 Db 35 RGGGGGGRSTSNRRKSKSYGARCMRSRRKACMSWMTGMSCTGPNMSWTKRRGSSW 94
 QY 139 GTTACCCGCTGTGTTACCCATGAATGCGACATGACTGCAAAAGCATGCCCGCTTAATCT 199
 Db 95 STTKNRKRGSTYSYTSYRRCYSMMWMCMSCYRSRSMCMWMSGSGSMCTGNASRCW 154
 QY 199 G 199
 Db 155 R 155

RESULT 9
US-08-469-412A-8

Patent No. 5856125

GENERAL INFORMATION:

APPLICANT: mavrounasitis, George J.
Blair, Donald G.
APPLICANT: fisher, Robert J.
Beal Jr., Gregory J.
APPLICANT: Athanasiou, Meropi A.
APPLICANT: Sgouras, Dionysios N.
TITLE OF INVENTION: The ERF Genetic Locus and Its Products
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:

STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA

COMPUTER READABLE

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS
SOFTWARE: Patent In Polystyrene

CURRENT APPLICATION DATA:
APPLICATION NUMBER: 706004111

FILING DATE: 05-JUN-1995
CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

REGISTRATION NUMBER: 37,330

TELECOMMUNICATION INFORMATION: 015280-229000

TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 332 base pairs

STRANDEDNESS: single

MOLECULE TYPE: DNA (genomic)

NAME/KEY: -

LOCALION: 1..332
OTHER INFORMATION

Some promoter region

Query Match

Matches	99;	Conservative	0;	Mismatches	111
---------	-----	--------------	----	------------	-----

47 GCAGTGCTC

[illegible]

-----GACAG1CGCCGCCGGG 179

107 GCATTCCTCCCCCCCCCA

Db 180 GGGCGTGGGCTGCCGGGCCCCGAGTGCATATTAAACCGGAGCGCGCGGGAGGG 239
Qy 167 CACATGCACTGCAAAAGCGATGCCCGTAAATCTGAAACGACATTACCCAGAGCGTTATG 226
Db 240 GAGAGGCTCTGAGAGCGGAGCGCGGTGTAGCGCGAGAGCGCGCCGACGCGCGCGGAC 299
Qy 227 GTAACGGCGCCGATGTAGGCGCAGAGTGGCGG 256
Db 300 GGGACGGGCGACGAGGCGCGCGGAGCCG 329

RESULT 10
US-09-021-715-8
; Sequence 8, Application US/09021715
; Patent No. 6194547
; GENERAL INFORMATION:
; APPLICANT: Mavrothalassitis, George J.
; Blair, Donald G.
; Fisher, Robert J.
; Beal Jr., Gregory J.
; Athanasiou, Meropi A.
; Sgouras, Dionysios N.
; TITLE OF INVENTION: The ERF Genetic Locus and Its Products
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/021,715
; FILING DATE: 10-Feb-1998
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Garrett-Mackowski, Eugenia
; REGISTRATION NUMBER: 37,330
; REFERENCE/DOCKET NUMBER: 015280-229000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0300
; TELEFAX: (415) 576-0200
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 332 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY:
; LOCATION: 1..332
; OTHER INFORMATION: /note="human ERF gene promoter region"
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-021-715-8

Query Match 7.1%; Score 32.4; DB 3; Length 332;
Best Local Similarity 47.1%; Pred. No. 0.53;
Matches 99; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 47 GCAGTGCCTCGGCTCGCTCGTTCCACATGAGGCGCGCGTAAATCAATAACGCGCGG 106
Db 120 GCGGGAGAGGGGCGCGGCGCATGTCTCCATGCGGAGCGCGCGGTACCTCGCGCGGG 179
Qy 107 GCATATGTTCCGGCCCGGACATATGATCAAGCTGTTACCCGCTGTTGTTAACCATGAATAG 166
Db 180 GGGCGTGGGCTCTCCGCGCCCGGAGTGTGATTAACCGGAGGCGCGCGCGGAGGG 239

Qy 167 CACATGCACTGCAAAAGCGATGCCCGTAAATCTGAAACGACATTACCGAGCGGTTATG 226
Db 240 GAGAGGCTCTGAGAGCGGAGCGCGGTGTAGCGCGAGAGCGCGCCGACGCGCGGAC 299
Qy 227 GTAACGGCGCCGATGTAGGCGCAGAGTGGCGG 256
Db 300 GGGACGGGCGACGAGGCGCGCGGAGCCG 329

RESULT 11
US-09-643-597-153/c
; Sequence 153, Application US/09643597
; Patent No. 6426072
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Kalos, Michael D.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Hoeken, Nancy
; APPLICANT: Fanger, Gary R.
; APPLICANT: Li, Samuel X.
; APPLICANT: Wang, Aljun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Henderson, Robert A.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C11
; CURRENT APPLICATION NUMBER: US/09/643,597
; CURRENT FILING DATE: 2000-08-21
; NUMBER OF SEQ ID NOS: 369
; SOFTWARE: PaStSeq for Windows Version 3.0
; SEQ ID NO 153
; LENGTH: 2007
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-643-597-153

Query Match 6.9%; Score 31.4; DB 4; Length 2007;
Best Local Similarity 46.5%; Pred. No. 2.8;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

Qy 119 GCCCGACTAGATGACGTGGTTACCGGTGTTTAAACCATGAATGACATGACATGC 178
Db 1337 GCTGTTCATTGAACACATGTTCCCGAATAATCTTCAAAAGATTGATTTT 1378
Qy 179 AAAGCGATGCCCGTAAATCTGAAACGACATTACCCAGAGCGGTTATGTAAAGCGCGG 238
Db 1377 AGTTAAAGTACTTAAAGTCACTCGGAAATTTTGCAATGAGTGAAGAACTGAGACT 1318
Qy 239 ATGTAGGCCAGGCGGTGGATTAATGTAATGTAAGTGAAGTCAAGATGCTTTCAAAATA 298
Db 1317 GCTTAGGGGGAATTTACGATTAAGGCTTGAATTAATCAACCGAGAAATCCAGTATA 1258
Qy 299 ATGCCACATGACGACGTGGAAGCTTAAACCTCGA 335
Db 1257 ATCTGCCCTCCCATTAAGTGTAGGAGAGAAAGCTCAA 1221

RESULT 12
US-09-480-884A-153/c
; Sequence 153, Application US/09480884A
; Patent No. 6482597
; GENERAL INFORMATION:
; APPLICANT: Wang, Tongtong
; APPLICANT: Fan, Liqun
; APPLICANT: Hoeken, Nancy A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C6
; CURRENT APPLICATION NUMBER: US/09/480,884A

/ CURRENT FILING DATE: 2001-08-27
/ NUMBER OF SEQ ID NOS: 330
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 153
/ LENGTH: 2007
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-480-884A-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGACTATGATCAGCTGTTACCCGCTGTTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGGAATTTTTCATGCAATGCAAGAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTGAATCACTCAGAAATGTTTCAAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGCATTTAGGCTTGAATTACTCAACCCAGAGAAAATCCAGTATA 1258
QY 299 ATGCCACATCGACCACTGGAAGCTTAAAACTCGCA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 13

US-09-542-615A-153/c
/ Sequence 153, Application US/09542615A
/ Patent No. 6518256
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Fan, Liqun
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Hosken, Nancy A.
/ APPLICANT: Fanger, Gary R.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
/ FILE REFERENCE: 210121.455C8
/ CURRENT APPLICATION NUMBER: US/09/542,615A
/ NUMBER OF SEQ ID NOS: 350
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 153
/ LENGTH: 2007
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-542-615A-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGACTATGATCAGCTGTTACCCGCTGTTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGGAATTTTGAATGCAAGTGAAGAAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTGAATCACTCAGAAATGTTTCAAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGCATTTAGGCTTGAATTACTCAACCCAGAGAAAATCCAGTATA 1258
QY 299 ATGCCACATCGACCACTGGAAGCTTAAAACTCGCA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 14
US-09-606-421B-153/c
/ Sequence 153, Application US/09606421B
/ Patent No. 6531315
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Fan, Liqun
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Hosken, Nancy
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Li, Samuel X.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A.W.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
/ FILE REFERENCE: 210121.455C9
/ CURRENT APPLICATION NUMBER: US/09/606,421B
/ NUMBER OF SEQ ID NOS: 358
/ SOFTWARE: FastSeq for Windows Version 3.0
/ SEQ ID NO: 153
/ LENGTH: 2007
/ TYPE: DNA
/ ORGANISM: Homo sapien
US-09-606-421B-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGACTATGATCAGCTGTTACCCGCTGTTTACCATGAATGCAATGCACCTGC 178
DB 1437 GCTTGTTCAATGAACACCATGTTCCCGAGAAAATCTCTACAAAGATGTTCAATTTT 1378
QY 179 AAGCGATGCCCGTAATCTGAAGACCATTAACCCAGAGCGGTTATGTTAGCGCCCG 238
DB 1377 AGTTAAAGTACTTAAAGTTCACTGGGAATTTTGAATGCAAGTGAAGAAACTGGAGCT 1318
QY 239 ATGTAGCCAGGGTGGGATTAATAGTACTATTGAATCACTCAGAAATGTTTCAAGAAATA 298
DB 1317 GCCTAGGGGAATTTTCAGCATTTAGGCTTGAATTACTCAACCCAGAGAAAATCCAGTATA 1258
QY 299 ATGCCACATCGACCACTGGAAGCTTAAAACTCGCA 335
DB 1257 ATCTGCCCTCCCATTTAGTGTAGAGAGAGAGCTCCAA 1221

RESULT 15
US-09-221-107-153/c
/ Sequence 153, Application US/09221107
/ Patent No. 6660838
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Tonglong
/ APPLICANT: Fan, Liqun
/ APPLICANT: Kalos, Michael D.
/ APPLICANT: Bangur, Chaitanya S.
/ APPLICANT: Hosken, Nancy
/ APPLICANT: Fanger, Gary R.
/ APPLICANT: Li, Samuel X.
/ APPLICANT: Wang, Aijun
/ APPLICANT: Skeiky, Yasir A.W.
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY OF LUNG CANCER
/ FILE REFERENCE: 210121.455C2
/ CURRENT APPLICATION NUMBER: US/09/221,107
/ NUMBER OF SEQ ID NOS: 161
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO: 153
/ LENGTH: 2007
/ TYPE: DNA
/ ORGANISM: Human
US-09-221-107-153

Query Match
Best Local Similarity 6.9%; Score 31.4; DB 4; Length 2007;
Matches 101; Conservative 0; Mismatches 116; Indels 0; Gaps 0;

QY 119 GCCCGACTATGATCAGCTGTTACCCGCTGTTTACCATGAATGCAATGCACCTGC 178

```

Db      1437 GCTTGTTCATTGACACCATGTTCCCGAAGAAATTCCTACAAAGATGTTCAATTTT 1378
Qy      179  AAAGCATGCCCGTAAATCTGAACGACCATTAACCGAGCGGTTATGTAACGGCGCG 238
Db      1377 AGTTAAAGTACTTAAAGTTCACTGCGAATTTTGCATGCAATGCAATGAGAAAAC 1318
Qy      239  ATGTAGGCGAGGCGGATTAATAGTACTATTGAATGACTGAGAAATGTTTCAAGAAATA 298
Db      1317 GCCTAGGGGAATTTCAAGATTAGGGCTTGAAATTACTCACCGAGAGAAATCCAGTATA 1258
Qy      299  ATGCCACCATCGACCACTGAAAGCTAAAGCTCCGA 335
Db      1257 ATCTGCCCTCCCATTAAGTGTAGAGAGAGCTCCA 1221

```

Search completed: March 16, 2004, 04:37:43
 Job time : 60.6647 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)
4179.927 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgsc.....ccacgcgtacacagatratraa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	36.4	8.0	2610	US-10-282-122A-30140	Sequence 30140, A
C 2	36.6	7.9	2667	US-10-282-122A-33185	Sequence 33185, A
C 3	35.8	7.9	954	US-10-282-122A-13780	Sequence 13780, A
C 4	34.2	7.5	1355	US-10-045-674-593	Sequence 593, App
C 5	34.2	7.5	2547	US-10-282-122A-36154	Sequence 36154, A
C 6	34	7.5	1878	US-10-369-493-43934	Sequence 43934, A
C 7	34	7.5	2886	US-09-801-368-131	Sequence 131, App
C 8	33.6	7.4	2301	US-09-815-242-4844	Sequence 4844, A
C 9	33.6	7.4	2301	US-09-815-242-9042	Sequence 9042, App
C 10	33.4	7.3	3140	US-10-425-114-29450	Sequence 29450, A
C 11	33.4	7.3	3485	US-10-424-599-64614	Sequence 64614, A
C 12	33.2	7.3	493	US-10-282-122A-26087	Sequence 26087, A
C 13	33.2	7.3	992	US-10-425-114-31747	Sequence 31747, A
C 14	32.8	7.2	568	US-10-333-631-1	Sequence 1, Appl1
C 15	32.8	7.2	718	US-10-027-632-26874	Sequence 26874, A

C 16	32.8	7.2	1059	US-09-738-626-482	Sequence 482, App
C 17	32.8	7.2	1113	US-09-738-626-481	Sequence 481, App
C 18	32.8	7.2	3309400	US-09-738-626-1	Sequence 1, Appl1
C 19	32.6	7.1	960	US-10-282-122A-11889	Sequence 11889, A
C 20	32.4	7.1	789	US-10-156-761-2022	Sequence 2022, App
C 21	32.4	7.1	9025608	US-10-156-761-1	Sequence 1, Appl1
C 22	32.2	7.1	972	US-10-333-631-4	Sequence 4, Appl1
C 23	32.2	7.1	1015	US-10-198-846-4377	Sequence 4377, App
C 24	32.2	7.1	1083	US-10-369-493-33133	Sequence 33133, A
C 25	32.2	7.1	4830	US-10-282-122A-27869	Sequence 27869, A
C 26	32	7.0	669	US-10-282-122A-31444	Sequence 31444, A
C 27	32	7.0	1001	US-10-425-114-14270	Sequence 14270, A
C 28	32	7.0	1152	US-10-156-761-303	Sequence 303, App
C 29	32	7.0	9025608	US-10-156-761-1	Sequence 1, Appl1
C 30	31.8	7.0	678	US-10-156-761-1219	Sequence 1219, App
C 31	31.8	7.0	930	US-09-815-242-6028	Sequence 6028, App
C 32	31.8	7.0	930	US-10-282-122A-20335	Sequence 20335, A
C 33	31.8	7.0	930	US-10-369-493-24480	Sequence 24480, A
C 34	31.8	7.0	978	US-10-369-493-24265	Sequence 24265, A
C 35	31.8	7.0	2766	US-10-282-122A-42042	Sequence 42042, A
C 36	31.6	6.9	951	US-10-369-493-43201	Sequence 43201, A
C 37	31.6	6.9	1338	US-10-369-493-37494	Sequence 37494, A
C 38	31.6	6.9	1642	US-09-887-576-469	Sequence 469, App
C 39	31.4	6.9	734	US-10-425-114-566	Sequence 566, App
C 40	31.4	6.9	1320	US-09-738-630-70	Sequence 70, Appl1
C 41	31.4	6.9	1467	US-10-282-122A-25701	Sequence 25701, A
C 42	31.4	6.9	1833	US-09-989-920-87	Sequence 87, Appl1
C 43	31.4	6.9	2007	US-09-735-705-153	Sequence 153, App
C 44	31.4	6.9	2007	US-09-850-716A-153	Sequence 153, App
C 45	31.4	6.9	2007	US-09-897-778-153	Sequence 153, App

ALIGNMENTS

RESULT 1
US-10-282-122A-30140/c
Sequence 30140, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haeselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
APPLICANT: Xu, H.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: EITRA.034A
CURRENT APPLICATION NUMBER: US/10/282, 122A
CURRENT FILING DATE: 2003-02-20
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230, 335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230, 347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267, 636

```

1 PRIOR FILING DATE: 2001-02-16
2 Remaining Prior Application data removed - See File Wrapper or PALM
3 NUMBER OF SEQ ID NOS: 78614
4 SOFTWARE: PatentIn version 3.1
5 SEQ ID NO 33185
6 LENGTH: 2667
7 TYPE: DNA
8 ORGANISM: Pseudomonas syringae
9 HS-10-282-122A-33185

```

```
Query Match      7.9%; Score 36; DB 12; Length 2667;
Best Local Similarity 43.7%;
Matches 159; Conservative 0; Mismatch 22
```

[illegible]

```

RESULT 3
US-10-282-122A-13780/c
; Sequence 13780, Application US/10282122A
; Publication No. US20040029129A1
GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl
; APPLICANT: Zyckind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELIPIRA.034A
CURRENT APPLICATION NUMBER: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-05-26

```

PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1
SEQ ID NO 13780
LENGTH: 954
TYPE: DNA
ORGANISM: Burkholderia mallei
US-10-282-122A-13780

Query Match 7.9%; Score 35.8; DB 12; Length 954;
Best Local Similarity 54.1%; Pred. No. 0.2;
Matches 73; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

Qy 55 CTGGCTGGCTGCTCCACATGAGGCGCGGCGGTAATCATTAACGGCGCGCAATAGT 114
Db 300 CTGGCGCCCATAGCGGAAGTAGGCGATGCGCGGTGATCCGGCGCGCGATGCGCGCTT 241
Qy 115 TCCGCGCCGACATGATGATGAGTGTGTTACCGGTGTTACCAATGAATGGCAATGCA 174
Db 240 GAGCGCATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181
Qy 175 CTGCAAGACGATGCC 189
Db 180 TTGCAGACGACGAGAC 166

RESULT 4

US-10-045-674-593
Sequence 593, Application US/10045674
Publication No. US2003023233A1
GENERAL INFORMATION:
APPLICANT: LADNER, ROBERT C.
APPLICANT: COHEN, EDWARD H.
APPLICANT: MASTRI, HORACIO G.
APPLICANT: ROONEY, KRISTIN L.
APPLICANT: HOEY, RENE
TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
TITLE OF INVENTION: LIBRARIES
FILE REFERENCE: DVAX/002 CIP2
CURRENT FILING DATE: US/10/045,674
PRIOR FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: 60/198,069
PRIOR FILING DATE: 2000-04-17
PRIOR APPLICATION NUMBER: 09/837,306
PRIOR FILING DATE: 2001-04-17
NUMBER OF SEQ ID NOS: 635
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 593
LENGTH: 1355
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: M3-III
FEATURE:
NAME/KEY: CDS
LOCATION: (1) : (1305)
US-10-045-674-593

Query Match 7.5%; Score 34.2; DB 15; Length 1355;
Best Local Similarity 26.9%; Pred. No. 0.86; Mismatches 169; Indels 0; Gaps 0;
Matches 90; Conservative 76; Mismatches 169; Indels 0; Gaps 0;

Qy 20 CAGATTGCGACGATCGTATGTTCTGCGAGTGTCTGCTGCGCTTCGTCACAAATGG 79
Db 755 CMCCKGTAAVAGCGKGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 814
Qy 80 GCGGCGCGGTAATCATTAACGGCGCGCAATAGTTCGCGCGCGCATATGATCACTGCG 139
Db 815 GYWSYGARGGWGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 874
Qy 140 TTACCGGTGTTTACCATGAAATGGACATGACACTGCAAGGATGCGCGTAATGTCG 199
Db 875 TGGCMAATGCTAAATGAAGGSCVATGACVRAAYGCGAAGABAAVGCRTCRARWSTG 934
Qy 200 AAGCACCATTACCCAGAGCGGTTATGTTAAGCGCGCGATGATAGCCAGGATGCGGATA 259
Db 935 AVGCYAAAGGYAAYTGWATVWSYGTGCGYACWGAATVAGTGTGCTGCTGATGATGATGAT 994
Qy 260 ATAGTACTATTAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 319
Db 995 TYGGYAGYGTWWSYGGYCTKCTCTAAAGGYAAGGWCYACVAGWATVYTCMGWGYTCA 1054
Qy 320 ACGTAAACTCCGATATTAATGCTGCGCGCAATAC 354
Db 1055 ATTCTCATATGCTCATGCTGGWAGYGGKAYAAH 1089

RESULT 5

US-10-282-122A-36154
Sequence 36154, Application US/10282122A
Publication No. US20040029129A1
GENERAL INFORMATION:
APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John
APPLICANT: Carr, Grant
APPLICANT: Yamamoto, Robert
APPLICANT: Forsyth, R.
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
FILE REFERENCE: ELITRA.034A
CURRENT FILING DATE: US/10/282,122A
CURRENT FILING DATE: 2003-02-20
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/230,335
PRIOR FILING DATE: 2000-09-06
PRIOR APPLICATION NUMBER: 60/230,347
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/267,636
PRIOR FILING DATE: 2001-02-09
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 78614
SOFTWARE: PatentIn version 3.1

SEQ ID NO 36154
 LENGTH: 2547
 TYPE: DNA
 ORGANISM: Streptococcus mutans
 US-10-282-122A-36154

Query Match
 Best Local Similarity 47.5%; Score 34.2; DB 12; Length 2547;
 Matches 102; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 133 CAGCTGTACCCGCTGTTGTTTACCATGAATGSCACATGCACTGGAAGCATCCCGT 192
 DB 850 CAGCAGGTGACACTGTTGTTGTCATGCACTGCGCCATCATAGTTGGTATCTGTGTA 909
 QY 193 AATCTGAAAGACCATTAACCCAGAGGGTAAAGTAAAGCGCCGATGAGCCAGGGT 252
 DB 910 ACTATGAAGTGGGAGTACCTTTGGCTGAATGAAGTTTGGCAATATGATGGAATAT 969
 QY 253 GCGGATATAGTACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 312
 DB 970 GTTGCTATGATGCTATTAACCAACTGGCATATTTTGAAGACTTTCAGACCAAGCGT 1029
 QY 313 CAGTGAACGCTAAACCTCCGATATTACTGTCGG 347
 DB 1030 GTTCTCATGCTTGAAGGCTATGCACTGACG 1064

RESULT 6

US-10-369-43934/c
 Sequence 43934, Application US/10369493
 Publication No. US20030233675A1
 GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 APPLICANT: Goldman, Barry S.
 APPLICANT: Chen, Xianfeng
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 FILE REFERENCE: 38-10(52052)B
 CURRENT APPLICATION NUMBER: US/10/369,493
 PRIOR FILING DATE: 2003-02-28
 PRIOR APPLICATION NUMBER: US 60/360,039
 NUMBER OF SEQ ID NOS: 47374
 SEQ ID NO 43934
 LENGTH: 1878
 TYPE: DNA
 ORGANISM: No. US20030233675A1loc punctiforme
 US-10-369-493-43934

Query Match
 Best Local Similarity 7.5%; Score 34; DB 15; Length 1878;
 Matches 115; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 96 TAAACGCGCGGCAATAGTCCGCGCGGACTAATGATGAGCTGTTACCCGTTGTTAC 155
 DB 553 TAAAGCGCACCGAATACTCTTCTTGTGTCGCGCGCAAGGTGATGATGAGCGGTTAC 494
 QY 156 CCATGAATGCGACATGCACTGCAAGCGATGCCGTAATCTGAAACGACCATTAACCA 215
 DB 493 CTGTAAATCCCAACACACACCGTTCGCAAAACCACTGACATGCAATGCTTACCCAG 434
 QY 216 GAGCGGTTAATGTAACGCGCGCGGATGAGCGCGGAGGTGGGATTAATGATCTATTGAAT 275
 DB 433 AGGGGTTGGCGACCATGCTCTTTGGCTTAAAGCGACGGAATTAATCTCTCTCAT 374
 QY 276 GACTCAGATGTTTCAAGAAATTAATGCACTGACGACGACGTAAGCAAGTAAACCTCGA 335
 DB 373 GTCCCGTCCACGCTTGAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 314
 QY 336 TATTACTGTC 345

DB 313 CAATACTGCC 304

RESULT 7

US-09-801-368-131
 Sequence 131, Application US/09801368
 Patent No. US20020128250A1
 GENERAL INFORMATION:
 APPLICANT: Busby, Robert
 APPLICANT: Call, Brian
 APPLICANT: Hecht, Peter
 APPLICANT: Holtzman, Doug
 APPLICANT: Madden, Kevin
 APPLICANT: Milne, Todd
 APPLICANT: Maxon, Mary
 APPLICANT: No. US20020128250A1man, Thea
 APPLICANT: Royer, John
 APPLICANT: Salama, Sofie
 APPLICANT: Sherman, Amir
 APPLICANT: Silva, Jeff
 APPLICANT: Summers, Eric
 TITLE OF INVENTION: Method for Improving Secondary Metabolite Production in Fungi
 FILE REFERENCE: 109272.147
 CURRENT APPLICATION NUMBER: US/09/801,368
 PRIOR FILING DATE: 2001-03-07
 PRIOR APPLICATION NUMBER: US 09/487,558
 PRIOR FILING DATE: 2000-01-19
 PRIOR APPLICATION NUMBER: US 60/160,587
 NUMBER OF SEQ ID NOS: 440
 SOFTWARE: Patent in version 3.0
 SEQ ID NO 131
 LENGTH: 2886
 TYPE: DNA
 ORGANISM: Saccharomyces cerevisiae
 US-09-801-368-131

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 2886;
 Matches 82; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 205 ACCATTACCGAGCGGTTATGTTAAGCGCGGATGAGCGGCGGATTAATAGT 264
 DB 1501 AACACACACATACATACATACATACATACATACATACATACATACATACATACAT 1560
 QY 265 ACTATTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 324
 DB 1561 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1620
 QY 325 AAAAATCCGATTAATGCTGCGGCAATACGCGGATTAATAC 366
 DB 1621 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1662

RESULT 8

US-09-815-242-4844
 Sequence 4844, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyckind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4844
LENGTH: 2301
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-815-242-4844

Query Match 7.4%; Score 33.6; DB 9; Length 2301;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 159 TGAATGGCAGATGCAATGCAAGGATCCCGTAATCTGAAAGACATTACCCAGAG 218
Db 1146 TGAAGGCTATATGCAAGGCTTATGATGTCAGCAATTTGAAAGTGACGATTAAT 1205
Qy 219 CGTTATGATGTAAGCGCCGATGTAAGCGGATGCGATTAATGATTAATGAACTGAC 278
Db 1206 GAGTTTATGATTTGATTAATGATGATGATTTCTTTAACTTTAATTTAGACCAAG 1265
Qy 279 TCAGATGCTTTGCAAAATTAATGCCACATGACGATGAAACGCTAAACCTCGATAT 338
Db 1266 TAAAAATTAATTAATTAATTAATGAAGGATTAATGATTAATGATTAATGATTAAT 1325
Qy 339 TACTGTGCGCAATACCGCGGTAATTAAC 366
Db 1326 TCAAAACACCTTTTGGATTAATTAAC 1353

RESULT 9

US-09-815-242-9042
Sequence 9042, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyekind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9042
LENGTH: 2301
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(2301)
US-09-815-242-9042

Query Match 7.4%; Score 33.6; DB 9; Length 2301;
Best Local Similarity 47.6%; Pred. No. 1.8;
Matches 99; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 159 TGAATGGCAGATGCAATGCAAGGATCCCGTAATCTGAAAGACATTACCCAGAG 218
Db 1146 TGAAGGCTATATGCAAGGCTTATGATGTCAGCAATTTGAAAGTGACGATTAAT 1205
Qy 219 CGTTATGATGTAAGCGCCGATGTAAGCGGATGCGATTAATGATTAATGAACTGAC 278
Db 1206 GAGTTTATGATTTGATTAATGATGATGATTTCTTTAACTTTAATTTAGACCAAG 1265
Qy 279 TCAGATGCTTTGCAAAATTAATGCCACATGACGATGAAACGCTAAACCTCGATAT 338
Db 1266 TAAAAATTAATTAATTAATTAATGAAGGATTAATGATTAATGATTAATGATTAAT 1325
Qy 339 TACTGTGCGCAATACCGCGGTAATTAAC 366
Db 1326 TCAAAACACCTTTTGGATTAATTAAC 1353

RESULT 10

US-10-425-114-29450
Sequence 29450, Application US/10425114
Publication No. US20040034888A1
GENERAL INFORMATION:
APPLICANT: Liu, Jingdong
APPLICANT: Zhou, Yinhua
APPLICANT: Kovalic, David K.
APPLICANT: Screen, Steven E.
APPLICANT: Tabaska, Jack E.
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(5313)B
CURRENT APPLICATION NUMBER: US/10/425,114
PRIOR FILING DATE: 2003-04-28
NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 29450
LENGTH: 3140
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: UC-GWFLMINSOY044G01_FLI
US-10-425-114-29450

Query Match 7.3%; Score 33.4; DB 12; Length 3140;
Best Local Similarity 45.8%; Pred. No. 2.6;
Matches 115; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

Qy 91 AATCATACGCGCGGCAATAGTCCGCGCCGACTATATGCTGCTTACCGGTT 150
Db 1801 AATCAAGGCAAGGCGCTTAATAGGCTTAATCAAGGATTAATGAAGCAATAGCTGTAA 1860
Qy 151 GTTACCATGAATTAATGACATGCACTGCAAGGATCCGCTAATCTGAAAGACAT 210
Db 1861 ACTGCATCTGAGAGGCTTAAGAAATGACAGGATCTCAAGATGTTGAAACAAAGTGT 1920
Qy 211 ACCAGAGCGGTTATGTAACGCGCGGATGATGAGGATGCGGATTAATGATTAAT 270
Db 1921 TCAGATGATGTTCTGCAAAAGGAAAGATGTTGATTAATTAATGATTAATGAGACTT 1980

TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments)
6103.863 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....caccagctaacagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
EST:
1: em_eacba:*
2: em_eacba:*
3: em_eacba:*
4: em_eacba:*
5: em_eacba:*
6: em_eacba:*
7: em_eacba:*
8: em_eacba:*
9: gb_eacba:*
10: gb_eacba:*
11: gb_eacba:*
12: gb_eacba:*
13: gb_eacba:*
14: gb_eacba:*
15: em_eacba:*
16: em_eacba:*
17: em_g88_hum:*
18: em_g88_inv:*
19: em_g88_pln:*
20: em_g88_vrt:*
21: em_g88_fun:*
22: em_g88_mam:*
23: em_g88_mus:*
24: em_g88_pro:*
25: em_g88_rtd:*
26: em_g88_pdg:*
27: em_g88_vrt:*
28: gb_g881:*
29: gb_g882:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match Length	ID	Description
C 1	232.6	51.0	680 12	Bu618688 Bu618688
C 2	37.4	8.2	484 9	AVA30994 AVA30994
C 3	37	8.1	590 14	CF848025 CF848025
C 4	37	8.1	610 14	CF842541 CF842541

Result No.	Score	Match Length	ID	Description
C 5	36.6	8.0	891 29	CNS02189
C 6	36.4	8.0	787 28	BZ568375
C 7	35.2	7.7	417 29	FR0020317
C 8	35.2	7.7	811 28	BZ469000
C 9	35.2	7.7	813 28	BZ204853
C 10	35	7.7	490 14	CAS88919
C 11	34.8	7.6	480 9	AU085223
C 12	34.6	7.6	406 9	AU241511
C 13	34.6	7.6	499 12	BJ006097
C 14	34.6	7.6	561 12	BJ001420
C 15	34.6	7.6	640 12	BJ000263
C 16	34.6	7.6	656 12	BJ004146
C 17	34.6	7.6	662 12	BJ010653
C 18	34.6	7.6	662 12	BJ011743
C 19	34.6	7.6	671 12	BJ012818
C 20	34.6	7.6	682 14	CF402142
C 21	34.6	7.6	691 12	BJ000180
C 22	34.6	7.6	702 12	BG441624
C 23	34.6	7.6	937 12	BG444255
C 24	34.4	7.5	365 9	AA746477
C 25	34.4	7.5	500 14	CA711477
C 26	34.4	7.5	827 28	BH438448
C 27	34.2	7.5	296 9	AV058630
C 28	34.2	7.5	523 12	BJ334624
C 29	34	7.5	588 14	CA731849
C 30	33.8	7.4	522 14	CA707989
C 31	33.8	7.4	625 13	BU882267
C 32	33.8	7.4	1098 29	CNS02GQP
C 33	33.6	7.4	584 28	AQ398232
C 34	33.6	7.4	594 12	BJ335653
C 35	33.6	7.4	653 12	BJ336964
C 36	33.6	7.4	712 14	CF437982
C 37	33.6	7.4	866 29	CNS02C24
C 38	33.6	7.4	896 28	CC33807
C 39	33.6	7.4	910 14	CA266034
C 40	33.4	7.3	525 9	AU242732
C 41	33.4	7.3	596 9	AU177600
C 42	33.4	7.3	649 12	BI261742
C 43	33.4	7.3	679 29	CC961526
C 44	33.4	7.3	695 28	BH63951
C 45	33.4	7.3	700 28	BH651371

ALIGNMENTS

RESULT 1
LOCUS Bu618688 680 bp mRNA linear EST 01-OCT-2003
DEFINITION Bu618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis CDNA clone XL186d22 5', mRNA sequence.

ACCESSION Bu618688
VERSION Bu618688.1 GI:37256713
KEYWORDS EST.

SOURCE Xenopus laevis (African clawed frog)

ORGANISM Xenopus laevis

REFERENCE Bukaryota, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.

AUTHORS Kikayama, A., Terasaka, C., Mochii, M., Ueno, N., Shin-I, T. and
Kobara, Y.

TITLE Expressed genes in X. laevis embryo

JOURNAL Unpublished (2001)

CONTACT Tadasu Shin-I

Center For Genetic Resource Information

National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6855
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp
The information of this clone is available through the following URL.

http://xenopus.nibb.ac.jp.

FEATURES Location/Qualifiers

source 1..680
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="X1186D22"
 /tissue_type="whole embryo"
 /dev_stage="stage 10.5"
 /clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match

Best Local Similarity 51.0%; Score 232.6; DB 12; Length 680;
 Matches 316; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGGAGTGGCTGGCT 60
 DB 593 ATGAACCTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGGAGTGGCTGGCT 60
 QY 61 GGGGTGCTCCACAAATGGGGCGGGCGGGTATCATACGGGGCGCAATAGTTCGGC 120
 DB 533 GGTGTGTTCTCAGTAGCGGGCGGGTATCACGGGTGGTGGCGGTAATTAAGGGG 474
 QY 121 CCGGACATGATCAGTGGTTACCGGTGTGTATCCCATGAAATGGCAATGACCTGCA 180
 DB 473 CCAATTTGAGCTGAACATTTTACAGTAGCGGTGGCGGTAACTNTGCACTTCTGCA 414
 QY 181 AGCGATGCCCTGAATCTGAAACGACATTCACGAGCGGTATGTTAAAGCGCGCAT 240
 DB 413 ACTGATGCCCGGTAACTCTGACTTGAATTAACGAGATGGCGCGGTATGTTGCAAGT 354
 QY 241 GTAGGCCAGGGTGGGATTAATGATTAATGAATCTGAGATGGTTTCAGAATATAT 300
 DB 353 GTTGTGAGGCTCAGTAGTGAAGCTCAATGATCTGACCAAGTGGCTTCGTTAAGC 294
 QY 301 GCCACATCGACAGTGGAAAGCTTAAATCTCGATTAATCTGCGGCAATAGCGCGT 360
 DB 293 GCTACTCTTATCAGTGAACGCGCAAAATTCGAAATACGCTTAAACAGTTGCGTGGT 234
 QY 361 AATAACCCCGCGCTGGTATTCAGACCGCATCTGATTCAGAGTATGTCGTCAGATT 420
 DB 233 GCGAAGGCTCTCAGTTGACAGACTGATCTAATCTCCTCGTCAAGTACTCAGATT 174
 QY 421 GGTTCGGCAACAGCCGCGCTAACAGTATTA 456
 DB 173 GGCTTGTGAACACGCGACGCTCATGATCTA 138

RESULT 2
 AV430994/c

LOCUS AV430994 484 bp mRNA linear EST 23-AUG-2000
 DEFINITION AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone
 ACCESSION AV430994
 VERSION AV430994.1 GI:8586219
 KEYWORDS EST.
 SOURCE Porphyra yezoensis
 ORGANISM Porphyra yezoensis
 Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Bangiaceae;
 Porphyra

REFERENCE 1 (bases 1 to 484)
 AUTHORS Nakaio, T., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Takata, S.
 TITLE Generation of 10,154 expressed sequence tags from a leafy gametophyte of a marine red alga, Porphyra yezoensis
 JOURNAL DNA Res. 7, 223-227 (2000)
 MEDLINE 20363100
 PUBMED 10907854
 COMMENT Contact: Erika Asamizu
 The First Laboratory for Plant Gene Research
 Kazusa DNA Research Institute

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
 Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
 Location/Qualifiers

FEATURES source 1..484
 /organism="Porphyra yezoensis"
 /mol_type="mRNA"
 /strain="TU-1"
 /db_xref="taxon:2788"
 /clone="PL027e10.r"
 /clone_lib="Porphyra yezoensis TU-1"
 /note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match

Best Local Similarity 8.2%; Score 37.4; DB 9; Length 484;
 Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 CGTAAACTCCGATATTACTGTCCGCAATAGCGGTAATTAACGCGCTGTTAA 380
 DB 441 CCAAAACACAAAGTCCCGCGCTCAACAGCGCTCAATACGCTGCGCAG 382
 QY 381 TCAGACCGATCGATTCAGCGTATGTCGTCAGTGGTTTGGCAACAGCCA 439
 DB 381 GCCGAAGCTGTGATCCAAAGTACTGTGTCAGGTTGCGGAAAGGCGCTCCCGCA 323

RESULT 3

CF848025 590 bp mRNA linear EST 30-OCT-2003
 LOCUS CF848025
 DEFINITION pSMA003x114f USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation_sma Phytophthora sojae cDNA clone
 ACCESSION CF848025
 VERSION CF848025.1 GI:38063679
 KEYWORDS EST.
 SOURCE Phytophthora sojae
 ORGANISM Phytophthora sojae
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;
 Phytophthora

REFERENCE 1 (bases 1 to 590)
 AUTHORS Tyler, B.
 TITLE Tyler, B. Not Published
 JOURNAL Unpublished (2003)
 COMMENT Contact: Tyler B
 Tyler lab

ORIGIN

1880 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7318
 Email: bmyler@vt.edu
 PCR Primers
 FORWARD: BK reverse primer
 BACKWARD: BK reverse primer
 Plate: 003 row: I column: 14
 Seq primer: BK reverse primer
 High quality sequence stop: 590.

FEATURES

source 1..590
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SMA003114"
 /tissue_type="mycelium"
 /cell_line="P6497"
 /dev_stage="mycelium"
 /lab_host="synthetic medium"
 /clone_lib="USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation_sma"
 /note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.1%; Score 37; DB 14; Length 590;
 Best Local Similarity 51.5%; Pred. No. 21;

Matches	85; Conservative	0; Mismatches	80; Indels	0; Gaps
Qy	26	TCGACGCAATGCTAGTCTTCTGCGAGTCTCTGCGTGGCGCTTCCACATGGGCGGCG	85	
Db	224	TTGTAGCGGTAGTCTCTTGTATGTCCGTGAGCTTGGGGAAGTCCCGCATGAAATGAGCCGCG	283	
Qy	86	GCGGTAATCATPAGCGCGCGGCATPAGTTCGGCGCCGAGCATGATACGCTTATCC	145	
Db	284	CCGTTGATGACACACAGCGCGCCACCAAGCGCGTGGCCGCGATATGCGCAGCGCTCG	343	
Qy	146	GTGTGTTTACCCCATGAAATGGCACATGCACTGCAAAAGCATGCC	190	
Db	344	ATCTTGAAGGCCAGCAGCGCGCATGATGAGCCCCACAGATGCC	388	
RESULT 4	CF842541			
LOCUS	CF842541			
DEFINITION	pefH020xg1.f USDA-IRAFS:Expression of Phytophthora sojae genes during infection and propagation_sHB Phytophthora sojae cDNA clone sHB020G21 5, mRNA sequence.			
ACCESSION	CF842541			
VERSION	CF842541.1			
KEYWORDS	EST.			
SOURCE	Phytophthora sojae			
ORGANISM	Phytophthora sojae			
REFERENCE	Eukaryota; Eriomanipiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.			
AUTHORS	1 (bases 1 to 610)			
TITLE	Tyler B.			
JOURNAL	Tyler B. Not Published			
COMMENT	Unpublished (2003)			
	Contact: Tyler B			
	Tyler lab			
	VBI			
	1880 Pratt Dr., Blacksburg, VA 24061, USA			
	Tel: 540-231-7318			
	Email: bmttyler@vt.edu			
	PCR Primers			
	FORWARD: BK reverse primer			
	BACKWARD: BK reverse primer			
	Plate: 020 row: G column: 21			
	Seq primer: BK reverse primer			
	High quality sequence stop: 610.			
FEATURES	Location/Qualifiers			
source	1..610			
	/organism="Phytophthora sojae"			
	/mol_type="mRNA"			
	/db_xref="taxon:67593"			
	/clone="sHB020G21"			
	/tissue_type="mycelium"			
	/cell_line="P6497"			
	/dev_stage="48 hr. post infection stage"			
	/lab_host="Soybean plant"			
	/clone_id="USDA-IRAFS:Expression of Phytophthora sojae genes during infection and propagation_sHB"			
	/note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"			
ORIGIN				
Query Match	8.1%; Score 37; DB 14; Length 610;			
Best Local Similarity	51.5%; Pred. No. 22;			
Matches	85; Conservative 0; Mismatches 80; Indels 0; Gaps 0;			
Qy	26	TCGACGCAATGCTAGTCTTCTGCGAGTCTCTGCGTGGCGCTTCCACATGGGCGGCG	85	
Db	212	TTGTAGCGGTAGTCTCTTGTATGTCCGTGAGCTTGGGGAAGTCCCGCATGAAATGAGCCGCG	271	
Qy	86	GCGGTAATCATPAGCGCGCGGCATPAGTTCGGCGCCGAGCATGATACGCTTATCC	145	
Db	272	CCGTTGATGACACACAGCGCGCCACCAAGCGCGTGGCCGCGATATGCGCAGCGCTCG	331	
Qy	146	GTGTGTTTACCCCATGAAATGGCACATGCACTGCAAAAGCATGCC	190	

	DB	332 ATCTTGAAGGCCACGAGCGCCGCATATGAGCCCCCAGACGATGGCC	376
	RESULT 5	CNS02189/c	
	LOCUS	CNS02189	
	DEFINITION	Tetradodon nigroviridis genome survey sequence PUC-Or1 end of clone	
	ACCESSION	AL198594.1 GI:7836745	
	VERSION	ALI98594	
	KEYWORDS	GSS: genome survey sequence.	
	SOURCE	Tetradodon nigroviridis	
	ORGANISM	Tetradodon nigroviridis	
	REFERENCE	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brothier,P., Quetier,F., Saurin,W. and Weissenbach,J. Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence Nat. Genet. 25 (2), 235-238 (2000)	
	AUTHORS	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fitzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis Genome Res. 10 (7), 939-949 (2000)	
	JOURNAL	20359837	
	MEDLINE	10899143	
	PUBMED	3 (bases 1 to 891)	
	REFERENCE	Genoscope.	
	AUTHORS	Direct Submission	
	JOURNAL	Submitted (12-APR-2000) GENOSCOPE - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequefigenoscope.cns.fr	
	COMMENT	- Web : www.genoscope.cns.fr This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetradodon.	
	FEATURES	Location/Qualifiers	
	SOURCE	1..891	
	ORIGIN	/organism="Tetradodon nigroviridis" /mol_type="genomic DNA" /db_xref="taxon:99883" /clone="141G11" /clone_lib="G" /note="Genoscope sequence ID : CONG141AD06SP1-end : PUC-Or1"	
	Query Match	8.0% Score 36.6; DB 29; Length 891;	
	Best Local Similarity	51.2%; Pred. No. 33;	
	Matches	66; Conservative 7; Mismatches 56; Indels 0; Gaps 0;	
OY	51 TGCTCTGGCTGGCTCGTTCCACAATGAGGCGCGCGGTATCATTAACGCGCGCGCAA	110	
Dd	437 TGTCTTCGCGCGGCGGTAAAGGAGGAGGATGAGCGCGGTAAACGCGTAGAGAGGCGW	378	
OY	111 TAGTTCGCGCGGACATGATGATGACTGCTTGTAACCGGTTGTTAACCATAATGCGACA	170	
Dd	377 TCTTGCAGGCGGTGTGWTATGTCAGCTGCTCMTCAWGTTGTCTTATTTCGCGCGAGC	318	
OY	171 TGCACTGCA 179		
Dd	317 CGCTTCGCA 309		

RESULT 6
BZ568375
LOCUS 787 bp DNA linear GSS 17-DEC-2002
DEFINITION pac62-164_7503.xl pac62-164 Pseudomonas aeruginosa genomic clone
ACCESSION BZ568375
VERSION BZ568375.1 GI:27201285
KEYWORDS GSS.
SOURCE Pseudomonas aeruginosa
ORGANISM Pseudomonas aeruginosa
REFERENCE Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas.
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) in press
COMMENT Contact: Chris K. Raymond
 Genome Center
 University of Washington
 Box 352145, Seattle, WA 98105-2145, USA
 Tel: 2062216954
 Fax: 2066857244
 Email: craymond@u.washington.edu
FEATURES
 source
 Location/Qualifiers
 1..787
 /organism="Pseudomonas aeruginosa"
 /mol_type="genomic DNA"
 /strain="2-164"
 /db_xref="taxon:287"
 /clone="pac62-164_7503"
 /clone_lib="pac62-164"
 /note="clinical isolate 2-164 Whole genomic shotgun library."
ORIGIN
 Query Match 8.0%; Score 36.4; DB 28; Length 787;
 Best Local Similarity 47.4%; Pred. No. 36;
 Matches 109; Conservative 0; Mismatches 121; Indels 0; Gaps 0;
 Db 17 TGGAGATTCGACGACGATGTTCTGCGAGTGTGCGCTGCGCTTCACAAAT 76
 146 TTGGCGGCTGCGCGACGTCGTGGAAGCGTCGACATCAGAGGTGTCTTACGCTG 205
 77 GGGCGCGCGCGCGATATCATACGCGCGCGCAATAGTTCCGCCGACTATGATCAGC 136
 206 TCGCGCGCGGTGTGGCCAGTTCCGCGTCGTCCAGCACTGACACGAGGTGGCGATGTTG 265
 137 TGGTACCGGTGTGTATCCCATGAATGACATGCACTGCAAGAGGATGCCCGTAAT 196
 266 TAGCCGCTGTCATGTGTCCAGCGCTCGCGGACGTTGCTTGTGATAGGGGGAG 325
 197 CTGAACGACCATTAACCGAGCGGTATGTGTACGGCGCGCGATGTAGGC 246
 326 GTGGCTTGGCCCTTGACGATGCGGAAAGAACCGGCTTGAAGTGGC 375
RESULT 7
FR0020317
LOCUS 417 bp DNA linear GSS 10-DEC-1997
DEFINITION F.rubripes GSS sequence, clone 040D20BD10, genomic survey sequence.
ACCESSION AL013202
VERSION AL013202.1 GI:2679570
KEYWORDS GSS; genome survey sequence.
SOURCE Takifugu rubripes (Fugu rubripes)
ORGANISM Takifugu rubripes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthopterygii; Acanthopterygii; Percomorphi; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Takifugu.
REFERENCE 1 (bases 1 to 417)
AUTHORS Elgar,G., Clark,M., Smith,S., Meek,S., Warner,S., Umrania,Y., Williams,G. and Brenner,S.
TITLE Direct Substitution
JOURNAL Submitted (08-DEC-1997) MRC Human Genome Mapping Project Resource Centre Hinxton, Cambridge, CB10 1SB. Email: b10helphgmp.mrc.ac.uk
COMMENT Vector: phagescript II KS
 V type: phagemid
 PRIMER: KS
DESCR: One pass dye-terminator sequencing of cosmid cloned genomic sequence.
FEATURES
 source
 Location/Qualifiers
 1..417
 /organism="Takifugu rubripes"
 /mol_type="genomic DNA"
 /db_xref="taxon:31033"
 /clone="040D20BD10"
 /clone_lib="cosmid 040D20"
ORIGIN
 Query Match 7.7%; Score 35.2; DB 29; Length 417;
 Best Local Similarity 50.3%; Pred. No. 59;
 Matches 76; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 Db 170 ATGCACTGCAAGACGATGCCGTAATCTGAAGACACATTACCCAGCGTTATGTA 229
 16 ATGCANCCAAAAGACACACNCNANTCCGACGCCGCAAGCGCGTTACGAGA 75
 230 ACCGCGCCGATGTAGCCAGGTCGATTAATGACTTATGTAAGTCACTGATGATGTT 289
 76 ATTGCGCTGTGTTCACAGCGCTGTGTTCATTGTAAGTCCAGAGCTTAATACNTG 135
 290 TGAAGAAATATGCCACCATGACCAAGTGA 320
 136 AGAGAACGATTATACACGACCTNAGAA 166
RESULT 8
BZ469000
LOCUS 811 bp DNA linear GSS 13-DEC-2002
DEFINITION BOOA058TF_BO_1.6_2_KB_tot Brassica oleracea genomic clone BOOA058, genomic survey sequence.
ACCESSION BZ469000
VERSION BZ469000.1 GI:26764546
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eustosids II; Brassicales; Brassicaceae; Brassica.
REFERENCE 1 (bases 1 to 811)
AUTHORS Town,C.D., Van Aken,S., Uteback,T., Koo,H. and Fraser,C.M.
TITLE Whole genome shotgun sequencing of Brassica oleracea
JOURNAL Unpublished (2001)
COMMENT Other GSSs: BOOA058TF
 Contact: Chris Town
 TIGR
 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TF
 Class: sheared ends.
FEATURES
 source
 Location/Qualifiers
 1..811
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOOA058"
 /clone_lib="BO_1.6_2_KB_tot"

/note="Vector: PHOS1; Site 1: BstXI; 1.6-2 kb sheared
total DNA inserted into PHOS1 using BstXI linkers"

ORIGIN

Query Match 7.7%; Score 35.2; DB 28; Length 811;
Best Local Similarity 47.7%; Pred. No. 78;
Matches 103; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

QY 146 GTGTTGTAACCATTAATGACATGACATGCAAGGATCCCGTAATTTGAAAGCA 205
DB 640 GTATTGAAATTAAGAGATTGTGAAGCTTGATTTGGCAAGCATTTGTAACCTGTGTTTC 581
QY 206 CCATTACCCAGAGCGGTTATGTAACGGCGCATGTAGCCAGGATGGCATTAATAGTA 265
DB 580 CCATTTCATCTACTGTTTATGCAAAATGCTTTGATCTAATTCATCTGATGTTTGAGCTG 521
QY 266 CTATTGAACTGACATGCAAAATGCTTTGCAAAATTAATGCAACCATGCAACGATGAAAGCTGA 325
DB 520 CTCCTTGCTGTCTAGGATAGTTTAAATCTTTGTCATTCATTAATGAAATGCA 461
QY 326 AAAATCCGATTAATCTGTGCGCCAAATAGCGCGGTA 361
DB 460 AATACACCTGATTAACACTTATTCAACCAAGATA 425

RESULT 9

BZ204853

LOCUS BZ204853 813 bp DNA linear GSS 11-OCT-2002
DEFINITION CH230-390D21.TV CHORI-230 Segment 2 Rattus norvegicus genomic clone
CH230-390D21, genomic survey sequence.

ACCESSION

BZ204853

VERSION BZ204853.1 GI:23862905

KEYWORDS

GSS.

SOURCE

Rattus norvegicus (Norway rat)

ORGANISM

Rattus norvegicus

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus;

1 (bases 1 to 813)

REFERENCE

AUTHORS

Zhao,S., Shetty,J., Shatsman,S., Tsengay,G., Geer,K.,

Shaartsbeyn,A., Gebregiorgis,E., Overton,L., Russell,D., Chen,D.,

Riggs,F., de Jong,P. and Fraser,C.M.

Rat BAC End Sequences from Library CHORI-230 MboI segment

Unpublished (1999)

Other_GSSs: CH230-390D21.TJ

COMMENT

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: szhao@igf.org

Clones are derived from the rat BAC library CHORI-230

(http://www.chori.org/bacpac/rat230.htm). For BAC library

availability, please contact Pieter de Jong (pje@jngmail.cho.org).

Clones may be purchased from BACPAC Resources

(http://www.chori.org/bacpac/or_ering_information.htm). BAC end

plate: http://www.chori.org/cdb/bac_ends/rat/bac_end_intro.html

Seq primer: T7

Class: BAC ends.

FEATURES

SOURCE

Location/Qualifiers

1..813

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/strain="BN/SNHed/MCW"

/db_xref="taxon:10116"

/clone="CH230-390D21"

/sex="Female"

/cell_type="Brain"

/clone_1ib="CHORI-230 Segment 2"

/note="Vector: PTARBAC1.3; Site 1: MboI; Site 2: MboI;

CHORI-230 Rat (BN/SNHed/MCW) BAC library produced by

ORIGIN

Pieter de Jong

Query Match 7.7%; Score 35.2; DB 28; Length 813;
Best Local Similarity 49.0%; Pred. No. 78;
Matches 94; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 178 CAAGCCATGCGCCGTAATCTGAAACGACATTAACCGAGCGGTTATGTAACGCGCC 237
DB 548 CAAGACCTCCACATCAAAACAGACACACTCAAACTAATTAAGAAAGAACTAGGAAACA 607
QY 238 GATGAGGCCAGGCGGTCGATTAATAGTACTATTGAACTGATCAGATGTTTGAAT 297
DB 608 TCTGGAACACATGCGGCACTGGAATAAATCTCTGAACAAACCAATGCTTATGCTCT 667
QY 298 AATGCCACATGACACCATGGAACGCTTAATAAATCTCCATTAATCTGCGCAATACGCG 357
DB 668 AAGATCAAGATCGACCAATGGATCTCATTAACGCAAAAGCTTCTGTAGGCAAGAGAC 727
QY 358 GGTAAATACGCC 369
DB 728 TGTGTTAGGAC 739

RESULT 10

CA588919/c

LOCUS CA588919/c 490 bp mRNA linear EST 19-NOV-2002
DEFINITION hab51c12.y1 Fugu UT7 adult skin Takifugu rubripes cDNA clone
IMAGE:6352126 5' similar to SW:CID4_MOUSE O35054 CLAUDIN-4 ; mRNA
sequence.

ACCESSION

CA588919

VERSION CA588919.1 GI:25133497

KEYWORDS

EST.

SOURCE

Takifugu rubripes (Fugu rubripes)

ORGANISM

Takifugu rubripes

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;

Tetraodontidae; Tetraodontidae; Takifugu.

1 (bases 1 to 490)

REFERENCE

AUTHORS

Clark,M., Johnson,S.L., Leirach,H., Lee,R., Li,F., Marra,M.,

Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers,Y.,

Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schuck,R.,

Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,

Waterston,R. and Wilson,R.

Washu Zebrafish EST Project 1998

Unpublished (1998)

COMMENT

Contact: Stephen L. Johnson

Washington University School of Medicine

444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: zbrafish@wustl.edu

Library materials provided by G. Elgar (UK MRC HGMP-RC) Library

and S. Metcabe (University of Tokyo, Institute of Medical Science

Laboratory of Aquatic Molecular Biology and Biotechnology) DNA

Sequencing by: Washington University Genome Sequencing Center Clone

distribution: I.M.A.G.E. Consortium/LNL, send email to:

info@image.jnl.gov

Seq primer: T3 ET from Amersham.

FEATURES

SOURCE

Location/Qualifiers

1..490

/organism="Takifugu rubripes"

/mol_type="mRNA"

/db_xref="taxon:31033"

/clone="IMAGE:6352126"

/sex="Female"

/tissue_type="skin"

/lab_host="DH10B (T1 phage-resistant)"

/clone_1ib="Fugu UT7 adult skin"

ORIGIN

[illegible]

VERSION AU085223.1 GT:9963206

Eukaryota; Viridiplantae; Streptophyta; Embryophyta: Tracheophyta.

YOSHIMURA, K., UGAWA, Y., YOSHIMARU, H., NAGASAKA, K.
and TSUMURA, Y.

PUBMED 11052197
CONFERENCE 1

FEATURES

Location/Qualifiers

Chlorine

latipes CDNA clone OLC53.01d similar to pir | T16763 | hypothetical

source
ORGANISM
Oryzias latipes (Japanese medaka)
Oryzias latipes

AUTHORS
Naruse, K., Mitani, H. and Tanaka, M.

Department of Biological Sciences
Graduate School of Science

Source	Location/Qualifiers
1	406

0; Gaps 0;

V
111 TAGTTCGCGCCAGCCTACCAAGCAGGAGGGG

Y 171 TGGACTGCA 179

DEFINITION BJ006097 MF01SSA cDNA *Oryzias latipes* cDNA clone MF01SSA08BD11 5', mRNA sequence.

ACCESSION BJ006097

VERSION BJ006097.1 GI:17359814

KEYWORDS EST.

SOURCE *Oryzias latipes* (Japanese medaka)

ORGANISM *Oryzias latipes*

REFERENCE Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H. Medaka EST Project in Takeda's lab Unpublished (2001)

AUTHORS Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.

TITLE Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.

FEATURES

source

1..499

/organism="Oryzias latipes"

/mol_type="mRNA"

/strain="Hd-r"

/db_xref="taxon:8090"

/clone="MF01SSA08BD11"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

/clone_1ib="MF01SSA cDNA"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 499; Best Local Similarity 54.3%; Pred. No. 94; Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

51 TGCCTGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 110

195 TGTCTGCGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 136

111 TAGTTCGGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 170

135 TCTTCAGCGCGGTGATCATGATCAGACAGCTCATGAGTTGCTGATCTCAGCGCAGC 76

171 TGCACTGCA 179

75 CGCTCTGCA 67

RESULT 14

BJ001420/c 561 bp mRNA linear EST 05-DEC-2001

LOCUS BJ001420 MF01SSA cDNA *Oryzias latipes* cDNA clone MF01SSA007C02 5', mRNA sequence.

ACCESSION BJ001420

VERSION BJ001420.1 GI:17361687

KEYWORDS EST.

SOURCE *Oryzias latipes* (Japanese medaka)

ORGANISM *Oryzias latipes*

REFERENCE Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H. Medaka EST Project in Takeda's lab Unpublished (2001)

AUTHORS Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.

TITLE Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics

1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.

FEATURES

source

1..561

/organism="Oryzias latipes"

/mol_type="mRNA"

/strain="Hd-r"

/db_xref="taxon:8090"

/clone="MF01SSA007C02"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

/clone_1ib="MF01SSA cDNA"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 561; Best Local Similarity 54.3%; Pred. No. 98; Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

51 TGCCTGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 110

230 TGTCTGCGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 171

111 TAGTTCGGCGCGCGCTGCTTCACAAATGGCGCGCGCGGTATCATTAACGGCGCGGCA 170

170 TCTTCAGCGCGGTGATCATGATCAGACAGCTCATGAGTTGCTGATCTCAGCGCAGC 111

171 TGCACTGCA 179

110 CGCTCTGCA 102

RESULT 15

BJ000263 640 bp mRNA linear EST 05-DEC-2001

LOCUS BJ000263 MF01SSA cDNA *Oryzias latipes* cDNA clone MF01SSA014A01 5', mRNA sequence.

ACCESSION BJ000263

VERSION BJ000263.1 GI:17358166

KEYWORDS EST.

SOURCE *Oryzias latipes* (Japanese medaka)

ORGANISM *Oryzias latipes*

REFERENCE Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H. Medaka EST Project in Takeda's lab Unpublished (2001)

AUTHORS Kohara, Y., Shin-i, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.

TITLE Medaka EST Project in Takeda's lab

JOURNAL Unpublished (2001)

COMMENT Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp.

FEATURES

source

1..640

/organism="Oryzias latipes"

/mol_type="mRNA"

/strain="Hd-r"

/db_xref="taxon:8090"

/clone="MF01SSA014A01"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

/clone_1ib="MF01SSA cDNA"

ORIGIN

Query Match 7.6%; Score 34.6; DB 12; Length 640;

Best Local Similarity 54.3%; Pred. No. 1e+02; Matches 70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY	51	TGCTCTGGCTGGCGCTCGTCCACAAATGGGGGGGGGGGATCATACGCGCGCGCA	110
Db	387	TGCTCTGGCGGGCGTACGGAGAGGGGATGACGGGGTGAACCGGAGACGAGGCGA	328
QY	111	TAGTTCGGCGCGGACTATGATGACTGTGTTACCGGTGTTGTTACCATGAATGGCACA	170
Db	327	TCTTGACGGCGTGTGATCATGATCAGCAGCTCCATCAGGTTGTCGTTGATCTCACCGCAGC	268
QY	171	TGCACTGCA	179
Db	267	CGCTCTGCA	259

Search completed: March 16, 2004, 04:29:12
Job time : 2235.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:49:25 ; Search time 1961.17 Seconds
(without alignments)
10077.856 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgcctaccagctatctaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_hcg: *
3: gb_in: *
4: gb_ov: *
5: gb_om: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vl: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_ov: *
22: em_ov: *
23: em_ph: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vl: *
30: em_hcg_hum: *
31: em_hcg_inv: *
32: em_hcg_other: *
33: em_hcg_mus: *
34: em_hcg_pln: *
35: em_hcg_rtd: *
36: em_hcg_mam: *
37: em_hcg_vrt: *
38: em_av: *
39: em_hcgo_hum: *
40: em_hcgo_mus: *
41: em_hcgo_other: *

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.8	90.5	456	6	I44909 Sequence 58
2	412.8	90.5	2067	1	U43280 Salmonella
3	404.8	88.8	5103	1	AJ002301 Salmonella
4	404.8	88.8	22411	1	AE008749 Salmonella
5	403.2	88.4	254050	1	AF627269 Salmonella
6	403.2	88.4	301983	1	AE016840 Salmonella
7	397.4	87.1	1048	1	STRGFB
8	288.6	63.3	361	6	I44908
9	280	61.4	2889	1	CSP515700
10	238.4	52.3	4680	1	ECCSGABDG
11	238.4	52.3	10346	1	AE000205
12	238.4	52.3	15047	1	D90741
13	237.2	52.0	2920	1	CPR515701
14	236.8	51.9	648	6	AX814811
15	236.8	51.9	648	1	ECOCSCGAA
16	228.6	50.1	1711	1	AF275733
17	228.6	50.1	10190	1	AE005315
18	228.6	50.1	327773	1	AP002554
19	227	49.8	306358	1	AE016759
20	195.2	42.8	2883	1	ESAS15702
21	185.8	40.7	230	1	SEUS53207
22	169.6	37.2	10370	1	AE015131
23	169.6	37.2	292504	1	AE016981
24	146.6	32.1	437	1	AF237726
25	93.2	20.4	19201	1	D90742
26	64.6	14.2	1212	1	EC0131756
27	48.2	10.6	78	6	AX814809
28	44.8	9.8	2000	6	AX814798
29	42	9.2	2000	6	AX655393
30	36.2	7.9	110243	2	AC146945
31	36.2	7.9	148860	8	OSJN00093
32	35.8	7.9	298900	1	AP005937
33	35.6	7.8	153232	9	AC024086
34	35.6	7.8	185124	9	AC009358
35	35.6	7.8	186208	9	AC142338
36	35.4	7.8	2915	14	PAD459805
37	35.4	7.8	2929	14	PAD554049
38	35.4	7.8	3300	14	FAD311719
39	35.4	7.8	3315	14	FAU26221
40	35	7.7	5014	3	PEPFRNA
41	35	7.7	12641	1	AE014731
42	35	7.7	349980	6	AX492785
43	35	7.7	349980	6	AX535952
44	34.8	7.6	301727	1	AE016960
45	34.8	7.6	347625	1	BX248356

ALIGNMENTS

RESULT 1
LOCUS I44909 456 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 58 from patent US 5635617.
ACCESSION I44909
VERSION I44909.1 GI:2469622
KEYWORDS
SOURCE
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 456)
AUTHORS Doran,J.L., Kay,W.W., Collinson,S.Karen, and Clouthier,S.C.
TITLE Methods and compositions comprising the agfa gene for detection of
Salmonella
JOURNAL Patent: US 5635617-A 58 03-JUN-1997;

FEATURES
Source location/Qualifiers
1..456
/organism="unknown"
/mol_type="unassigned DNA"

Query Match 90.5%; Score 412.8; DB 6; Length 456;
Best Local Similarity 94.1%; Pred. No. 6,6e-108;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

ORIGIN

```

QY 1 ATGAAGACTTTTAAAGTGCGACGATTCGACGAACTGAGTTCGCGAGCGCTGGCT 60
DB 1 ATGAAGACTTTTAAAGTGCGACGATTCGACGAACTGAGTTCGCGAGCGCTGGCT 60
QY 61 GCGCTGCTTCACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
DB 61 GCGCTGCTTCACATGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 120
QY 121 CCGGACTCAACGTTGAGCATTATCAGTACCGGTCGCGTAAAGCGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTATCAGTACCGGTCGCGTAAAGCGCTGCTGCTGCA 180
QY 181 AGCGATGCGCCGTAATATGATCAGTACCGGTCGCGTAAAGCGCTGCTGCTGCA 240
DB 181 AGCGATGCGCCGTAATATGATCAGTACCGGTCGCGTAAAGCGCTGCTGCTGCA 240
QY 241 GCGAGCGACGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 GCGAGCGACGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACGATGACGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 360
DB 301 GCCACGATGACGACGATGACGATGACGATGACGATGACGATGACGATGACGATGAC 360
QY 361 AATPACGCGCGCGTGTATATCAGACCGGATGATGATGATGATGATGATGATGAT 420
DB 361 AATPACGCGCGCGTGTATATCAGACCGGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTTTGGGACACACCGGCGGTAACCGATTTAA 456
DB 421 GGTTTGGGACACACCGGCGGTAACCGATTTAA 456

RESULT 2
SEU43280 2067 bp DNA linear BCT 14-FEB-1996
LOCUS Salmonella enteritidis agfBAC operon: fimbria-like protein
DEFINITION AgfC (agfC) genes, complete cds.
ACCESSION U43280
VERSION U43280.1 GI:1184712
KEYWORDS
SOURCE
ORGANISM Salmonella enteritidis
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
1 (bases 1 to 2067)
Doran, J.L., Collinson, S.K., Burian, J., Sarlos, G., Todd, E.C.,
Munro, C.K., Kay, C.M., Baner, P.A., Peterkin, P.I. and Kay, W.W.,
DNA-based diagnostic tests for Salmonella species targeting agfA,
the structural gene for thin, aggregative fimbriae
J. Clin. Microbiol. 31 (9), 2263-2273 (1993)
JOURNAL MEDLINE 94013373
PUBMED 8104955
AUTHORS 2 (bases 1 to 2067)
Collinson, S.K., Clouthier, S.C., Doran, J.L., Baner, P.A. and
Kay, W.W.
TITLE Salmonella enteritidis agfBAC operon encoding thin, aggregative
fimbriae
J. Bacteriol. 178 (3), 662-667 (1996)
JOURNAL MEDLINE 96146512
PUBMED 850497
REFERENCE 3 (bases 1 to 2067)

```

AUTHORS Collinson, S.K., Doran, J.L., Baner, P.A. and Kay, W.W.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-1995) S. Karen Collinson, Biochemistry and
Microbiology, University of Victoria, P.O. Box 1800 STN CSC,
Victoria, BC V8W 2P6, Canada
FEATURES
Source location/Qualifiers
1..2067
/organism="Salmonella enteritidis"
/mol_type="genomic DNA"
/strain="27655-3b"
/db_xref="taxon:592"
/map="between puta and pyrC"
571..576
598..603
696..1151
/gene="agfB"
/note="fimbria-like protein precursor"
/codon_start=1
/transl_table=11
/product="AgfB precursor"
/protein_id="AAC43598.1"
/db_xref="GI:1184713"
/translation="MKKLLPMLTLLGAPGATATYDARSEYFANVELSSFN
SISQAYGNSAIIQKSGNKANTTYGTQKTAAYVQKSHMAIRVQR"
696..752
/gene="agfB"
753..1148
/gene="agfB"
/product="AgfB"
/note="putative signal sequence of 21 amino acids;
fimbria-like protein"
1193..1648
/gene="agfA"
1193..1648
/gene="agfA"
/note="thin aggregative fimbria subunit precursor; major
fimbrial subunit of thin aggregative fimbriae precursor"
/codon_start=1
/transl_table=11
/product="AgfA fimbria precursor"
/protein_id="AAC43599.1"
/db_xref="GI:1184714"
/translation="MKLLKVAAPALIVGSGALAGVPMGGGNGHNGSGNSGPDST
LSIYQGSANALALQSDARKSRTTIGSGYGADVGAGANSTIELTQNGFRNNAT
IDQWNAKNSDITGVGGNNALVNOTASDSSVMRQVGFNNATANDY"
1292..1354
/gene="agfA"
1352..1645
/gene="agfA"
/note="thin aggregative fimbria subunit; major fimbrial
subunit of thin aggregative fimbriae"
1667..1696
1710..2036
/gene="agfC"
1710..2036
/gene="agfC"
/function="unknown"
/codon_start=1
/transl_table=11
/product="AgfC"
/protein_id="AAC43600.1"
/db_xref="GI:1184715"
/translation="MTLLALALSNQITFTTQGGDIYVIVPVTLNBPVCQVOITL
SVRDSVGQSHQKTLSPANQPIEHSRLSVISSEDSKILIVYSDQSLHLSQQ
WPSAQ"

ORIGIN

```

Query Match 90.5%; Score 412.8; DB 1; Length 2067;
Best Local Similarity 94.1%; Pred. No. 7.2e-108;

```

Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;			
Qy	1	ATGAACCTTTAAAGTGGACGACATTCGACGAATCGAGTTCTTGGCAGTCTTGCT	60
Db	1193	ATGAACCTTTAAAGTGGACGACATTCGACGAATCGAGTTCTTGGCAGTCTTGCT	1252
Qy	61	GGCGTCGTTCCACATAGGGCGGCGCGGTATCATTAACGGCGGCGCAATAGTCCGGC	120
Db	1253	GGCGTCGTTCCACATAGGGCGGCGCGGTATCATTAACGGCGGCGCAATAGTCCGGC	1312
Qy	121	CCGACATCAACGTTGAGCATTTATCATGACGTTCCGTTACGCTCGCTTCTTCGCA	180
Db	1213	CCGACATCAACGTTGAGCATTTATCATGACGTTCCGTTACGCTCGCTTCTTCGCA	1372
Qy	181	AGCGATGCCCGTAAATATGATCAGTGGTACCCGTGTTTACCATGAATGGCAAT	240
Db	1373	AGCGATGCCCGTAAATATGAAACGACATTCACCGACGGGTATGTACGGCCCGAT	1432
Qy	241	GCAGGCGCAGGTGCGGATATAGTACTATTGAACTGATCAGAAATGTTTCAGAAATAT	300
Db	1433	GTAAGCCAGGTTGCGGATATAGTACTATTGAACTGATCAGAAATGTTTCAGAAATAT	1492
Qy	301	GCCACCATCGACCACTGGAACCGTAAAACTCCGATATTACTGTGCGCAATACGGCGGT	360
Db	1493	GCCACCATCGACCACTGGAACCGTAAAACTCCGATATTACTGTGCGCAATACGGCGGT	1552
Qy	361	AATAACGGCGGCTGTTATCATGACCCGATCTGTTCACGCTAATGTCGCTCAGATT	420
Db	1553	AATAACGGCGGCTGTTATCATGACCCGATCTGTTCACGCTAATGTCGCTCAGATT	1612
Qy	421	GGTTTGGCAACAACGCGACGCTAACCATATTAA	456
Db	1613	GGTTTGGCAACAACGCGACGCTAACCATATTAA	1648
RESULT 3			
STAJ2301 5103 bp DNA linear BCT 15-NOV-2000			
LOCUS Salmoneilla typhimurium csgc, csgf, csgg, csgd, csgh, csgh, and e			
DEFINITION Salmoneilla typhimurium csgc, csgf, csgg, csgd, csgh, and e			
ACCESSION AJ002301			
VERSION AJ002301.1 GI:2739232			
KEYWORDS csgA gene; csgB gene; csgC gene; csgD gene; csgE gene; csgF gene;			
SOURCE Salmoneilla typhimurium			
ORGANISM Salmoneilla typhimurium			
REFERENCE 1 Enterobacteriaceae; Salmoneilla.			
AUTHORS Romling U., Bian Z., Hammar M., Sierralta W.D. and Normark S.			
TITLE Culi fibers are highly conserved between Salmoneilla typhimurium			
JOURNAL Submitted (29-OCT-1997) Romling U., Department of Bacteriology,			
Karolinska Institute, MTC, Box 280, Stockholm, S-17177, SWEDEN			
FEATURES			
SOURCE location/Qualifiers			
1. 5103			
/organism="Salmoneilla typhimurium"			
/mol_type="genomic DNA"			
/strain="SR-11"			
/db_xref="taxon:602"			
<1..332			
/note="ORF179a"			
/codon_start=2			
/transl_table=1			
/product="hypothetical protein"			
/protein_id="CAA05311.1"			
CDS			
10_signal			
-35_signal			
/db_xref="GI:2739233"			
/db_xref="GOA:O54290"			
/db_xref="SWISS-PROT:O54290"			
/translation="CTLSCGMVAVLVIIHGSALAPHEIYSVLTGIVAFIMCTIOAKQ			
LVLPFGVTFIGACATPAGGDMRLVPLSLGLIFGVMKNSGLMLASRREHSANT			
AVLK"			
complement (361..1194)			
/gene="csgg"			
complement (361..1194)			
/gene="csgd"			
/function="11poprotein"			
/codon_start=1			
/evidence=experimental			
/transl_table=1			
/product="Csgd protein"			
/protein_id="CAA05312.1"			
/db_xref="GI:2739234"			
/db_xref="GOA:O54291"			
/db_xref="SWISS-PROT:O54291"			
/translation="MPRLIVAVLLISGCLTAPKQAAKPTLMPRAQYKDLTHLPA			
PTGKIFVSVNIQDETQCFKPYPSANSTAVPOSATMLTALAKSRFTFLERQGLQ			
NLINERKIIRAAQENGTVAMNRIPLQSLTAAITMVGSIIGYSNVKSGVGARFG			
IGADTQYQLODIAVNLRVNVSTGELISSVTSKTLISYEQACVFFPIIDYORLLGE			
IGYSNRPVMLCLMSAIETGVIFLINDIGRLMDLQNKADRDNDILVKYRELVPPE			
S"			
complement (1221..1637)			
/gene="csgf"			
complement (1221..1637)			
/gene="csgh"			
/codon_start=1			
/transl_table=1			
/product="Csgf protein"			
/protein_id="CAA05313.1"			
/db_xref="GI:2739235"			
/db_xref="SWISS-PROT:O54292"			
/translation="MRVKAHVLLMLFSPLTWAGNMTFOPRNPNFGNPNNGSEPLNS			
AOAONSYKDPAYNDPFIETPSALNDPTOAIOSQILGGLLTNIWTKRGRVNTDFTI			
DIANRBDQLQNLVTDKRTGRSTREVSGLQTQSTDP"			
complement (1664..2059)			
/gene="csgs"			
complement (1664..2059)			
/gene="csgh"			
/codon_start=1			
/transl_table=1			
/product="Csgs protein"			
/protein_id="CAA05314.1"			
/db_xref="GI:2739236"			
/db_xref="SWISS-PROT:O54293"			
/translation="MKRYLTMIVAAELLPATGNLHANEVEVPGLLTDHTVSSIGHE			
FYAFSPKMSSEVTGNLTINERBSARGSWITTVNDVIFQTLFPMKDFEFTVVF			
ALMOTERALKNRQIDQTLSTSLARDE"			
complement (2064..2714)			
/gene="csgd"			
complement (2064..2714)			
/gene="csgh"			
/function="transcriptional regulator"			
/codon_start=1			
/transl_table=1			
/product="Csgd protein"			
/protein_id="CAA05315.1"			
/db_xref="GI:2739237"			
/db_xref="GOA:O54294"			
/db_xref="SWISS-PROT:O54294"			
/translation="MFEVSHSHGHTLLITKQSLQATALLQHLKQSLATITKLANIQ			
RSLDEISAGCIVLMDMADKLIHYQDMLSRNNNIKTLLNTPDPYRELENNP			
HINGEVPATEDOEHVVSGLQILGECVFQKSLASYITISGVRYVSTSLATLTHE			
KEILNKLRIGASNNNEIRSLPFIENSTVKTLYNLFKKIAVKNRTQAVSWANDNLRR"			
complement (2869..2896)			
/gene="csghDERG"			
complement (2869..2874)			
/gene="csghDERG"			
complement (2891..2896)			
/gene="csghDERG"			

```

gene
3344..3376
/gene="csGBA"
-35_signal
3344..3349
/gene="csGBA"
-10_signal
3371..3376
/gene="csGBA"
gene
3469..3924
/gene="csGB"
CDS
3469..3924
/gene="csGB"
/function="nucleator"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CsGB protein"
/protein_id="CA05316.1"
/db_xref="GI:2739238"
/db_xref="GOA:P55226"
/translacion="MKKLLFMWLTITIGAPGATATATNDLASEYNFAVNEISKSEFN
QAALIGVGTNDNARVROBSSKLSVTSQGGNNRKNVDQAGNPNFAYIEOTGNANDA
SISOSAYGNSAALIOKSGNKANTITGYGTAKTAVVQKSHMAIVTOR"
3469..3531
/gene="csGB"
3532..3921
/gene="csGB"
/product="CsGB protein"
/function="nucleator"
3966..4421
/gene="csGA"
3966..4421
/gene="csGA"
/function="fiber subunit"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="CsGA protein"
/protein_id="CA05317.1"
/db_xref="GI:2739239"
/db_xref="GOA:P55225"
/translacion="MKLKVAAPAIIVSGSALAGVYPOMGGGNNNGGNSGPDST
LSIYOGSANAALIOKSGNKANTITGYGTAKTAVVQKSHMAIVTOR"
3966..4025
/gene="csGA"
4026..4418
/gene="csGA"
/product="CsGA protein"
/function="fiber subunit"
4483..4809
/gene="csGC"
4483..4809
/gene="csGC"
/codon_start=1
/transl_table=11
/product="CsGC protein"
/protein_id="CA05318.1"
/db_xref="GI:2739240"
/db_xref="SWISS-PROT:P55227"
/translacion="MHTLLIALNSQITFTTQGGDIYTVIPQTLNPPCYOVQIT
SVRDVGOGSHTOOKTILIPANQPIELSRSLNISSEDSVKIITVPSDGSLSHLSQ
WPSAQ"
4483..4533
/gene="csGC"
4534..4806
/gene="csGC"
/product="CsGC"
4940..5103
/gene="ORF103"
4940..5103
/gene="ORF103"
/codon_start=1

```

```

ORIGIN
Query Match      88.8%; Score 404.8; DB 1; Length 5103;
Best Local Similarity 93.0%; Pred. No. 1.6e-105;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

/transl_table=11
/protein_id="CA05319.1"
/db_xref="GI:2739241"
/translacion="MSVYKKNIPAIIGLICAFPIHSAVGQTVQGVHFRGAIVEPL
CDISTHMENT"

1 ATGAACCTTTTAAAGTGGAGCATTCGCGCATGTAAGTTCTGCACTGCTGCT 60
3966 ATGAACCTTTTAAAGTGGAGCATTCGCGCATGTAAGTTCTGCACTGCTGCT 4025
61 GGCCTGCTTCACATATGGCGCGCGGATCATATACGGCGGCAATAGTCCCGGC 120
4026 GGCCTGCTTCACATATGGCGCGCGGATCATATACGGCGGCAATAGTCCCGGC 4085
121 CCGGACTCAGCTTGGCATTTATCATACGTTCCGTTACCGCTGCTCTGCA 180
4086 CCGGACTCAGCTTGGCATTTATCATACGTTCCGTTACCGCTGCTCTGCA 4145
181 AGCGATGCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCAAT 240
4146 AGCGATGCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGCAAT 4205
241 GCAGCCGAGGTCGCGATATATGATCAGCTGTTACCCGTTGTTACCATGAATGCAAT 300
4206 GTCAGCCGAGGTCGCGATATATGATCAGCTGTTACCCGTTGTTACCATGAATGCAAT 4265
301 GCCACCATGACAGTGAAGCTTAAATCTCCATATATCTGTCGCGCATATCGCGCT 360
4266 GCCACCATGACAGTGAAGCTTAAATCTCCATATATCTGTCGCGCATATCGCGCT 4325
361 AATAACGCGCGCTGTTATATGACGCGATCTGATTCACGCGTAATGTCGTCAGGTT 420
4326 AATAACGCGCGCTGTTATATGACGCGATCTGATTCACGCGTAATGTCGTCAGGTT 4385
421 GGTTTTGGCAACCAACGCGGCTTACCATGATTA 456
4386 GGTTTTGGCAACCAACGCGGCTTACCATGATTA 4421

RESULT 4
AE008749 22411 bp DNA linear BCT 23-APR-2003
LOCUS Salmonella typhimurium LT2, section 53 of 220 of the complete
DEFINITION genome.
ACCESSION AE008749 AE006468
VERSION AE008749.1 GI:16419641
KEYWORDS
SOURCE
ORGANISM
Salmonella typhimurium LT2
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
1 (bases 1 to 22411)
McClelland, M., Sanderson, K.E., Spieth, J., Clifton, S.W.,
Latterelle, P., Courtney, L., Porwollik, S., Ali, J., Dante, M., Du, F.,
Hou, S., Layman, D., Leonard, S., Nguyen, C., Scott, K., Holmes, A.,
Grewal, N., Mulvaney, E., Ryan, E., Sun, H., Flores, L., Miller, W.,
Stonking, T., Nhan, M., Waterston, R. and Wilson, R.K.
Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2
JOURNAL Nature 413 (6858), 852-856 (2001)
MEDLINE 21534948
PUBMED 11677609
REFERENCE
1 (bases 1 to 22411)
2 The Salmonella typhimurium Genome Sequencing Project
DIRECT SUBMISSION
TITLE Submitted (29-MAR-2001) Genome Sequencing Center, Department of
JOURNAL

```

COMMENT

Genetics, Washington University School of Medicine, 4444 Forest Park Boulevard, St. Louis, MO 63108, USA
 COMMENT Supported by NIH grant 5U 01 AI33283

Coding sequences below are predicted from manually evaluated computer analysis, using similarity information and the programs; GLIMMER; <http://www.tigr.org/softlab/glimmer/glimmer.html> and GeneMark; <http://opal.biology.gatech.edu/GeneMark/>

EC numbers were kindly provided by Junko Yabuzaki and the Kyoto Encyclopedia of Genes and Genomes; <http://www.genome.ad.jp/kegg/>, and Pedro Romero and Peter Karp at EcoCyc; <http://ecocyc.org/ecocyc/PangeaSystems.com/ecocyc/>

The analyses of ribosome binding sites and promoter binding sites were kindly provided by Heladia Salgado, Julio Collado-Vides and ReguonDB; http://kinich.cifn.unam.mx:8850/db/reguondb_intro.frameset

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistries or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one m3 subclone.

FEATURES

SOURCE

Location/Qualifiers

1..22411

/organism="Salmonella typhimurium LT2"

/mol_type="genomic DNA"

/strain="LT2; SGSC 1412; ATCC 700720"

/db_xref="ATCC:700720"

/note="LT2"

/gene="phoH"

/note="synonym: STM1126"

434..439

/gene="phoH"

/note="putative RBS for phoH; ReguonDB:STMS1H001398"

454..1308

/gene="phoH"

/note="similar to E. coli PhoB-dependent, ATP-binding pho regulon component; may be helicase; induced by P starvation (AAC74105.1); Blastp hit to AAC74105.1 (354 aa), 92% identity in aa 71 - 354"

/codon_start=1

/transl_table=1

/product="PhoB-dependent ATP-binding pho regulon component"

/protein_id="AAL20057.1"

/db_xref="GI:16419642"

/translation="MCANRVQENAGDVYQLKYLIMGKQAVIKARREARVLRDS RSHKQREESVTSLVQMGVEAIGWADSRDTPVKANEAAHLYNIDSKQIFAT GEACGKTVISAAKAALIKHDEVERITVTRPVLOADEDLGLPDIDAEKAPRPV YDVLKRLGASFMQCLRPETGKVIAPFAPRGTPENAVIILDEAQNVTAAAMKMP LTRGENTVTVNGDITOCDELPRGVRSLDALERFEDENVGIVHFKDKCVSALC ORTHAYS"

/complement(1414..2303)

/gene="STM1127"

/complement(1414..2295)

/note="similar to E. coli orf, hypothetical protein (AAC75480.1); Blastp hit to AAC75480.1 (285 aa), 23%

identity in aa 15 - 264"

/codon_start=1

/transl_table=1

/product="putative transcriptional regulator"

/protein_id="AAL20058.1"

/db_xref="GI:16419643"

/translation="MEQPRRLKPKGKILDTLGMOKSITRASQRIAYITAPQVTO SSIDLSRETQGEATVIRFCRTLGKGFQKFDKDLAELATTSDDSSPLDDEVSE SDDAAIGLKLQNTISNVLTSLNLIDQVLTGVADLRHCHSVYIFGVSGGITALD MKHKLMTIRGLGDAVSNHFMVQATLLKADVAMGVSHSGTSETVSHLRLAQAG

RBS

TTVAITNIGSPICEADFCLLINGRQGLQGSIGTKAQLFVPDLTYTLVYSSPE OARSKARTNALMTK"

/complement(2258..2303)

/gene="STM1127"

/note="putative RBS for STM1127; ReguonDB:STMS1H001399"

/complement(2580..4088)

/gene="STM1128"

/note="similar to E. coli putative cotransporter (AAC76102.1); Blastp hit to AAC76102.1 (571 aa), 22%

identity in aa 7 - 478"

/codon_start=1

/transl_table=1

/product="putative sodium/glucose cotransporter"

/protein_id="AAL20059.1"

/db_xref="GI:16419644"

/translation="MTHSFGIVNYLVLFGLAMLVGVYFSRROKTADYPRGGR VPGWAGVSPATLSTTTPMSIPAKAFTSDWTFILQVLAAILPLVYFYIPERK LKYSATBYLEARDVRCRLPASMPLFHIGRAITFTFLVALRPIAIDPVLV LKVSCTITVMGSGIEGVITWDVLOGLLSAFLIFVLCVKQGGIDEIFVTQQA DKFPAPQFHWSTESTVPMVLMIGFLPANIQQFASODVVRVYVTSIBETKTLT NAKLVAVIPVFPFAGSALFVYVOHQPLPAGNTGGLPLFVTEMPVGIAGIIA AIFPAOSISSISNSISCSFSDIYORLSHKRTPEMRKIAKLVLVAGIISAA VMIYMADESEIWDAPNSLIGMGPMGLTKGIFFRANAGSAVGIITIVITVGA RYATDINPFYGVIGSLSVITSGVIFAPLPAPPLTDERBKRTL"

/complement(4083..4088)

/gene="STM1128"

/note="putative RBS for STM1128; ReguonDB:STMS1H001400"

/complement(4413..5099)

/gene="STM1129"

/complement(4413..5093)

/gene="STM1129"

/note="similar to E. coli putative enzyme (AAC76255.1); Blastp hit to AAC76255.1 (229 aa), 70% identity in aa 1 - 226"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20060.1"

/db_xref="GI:16419645"

/translation="MSLLARLEQSVHENGGLIVSCOPVPSPMKPELVAAQMAAS AGAVAVLEGENTRTVPHLSVPIIGIKRDLGSPRTIPYODNALAQAQDIT AFDAFSRPRPDIDSLTRIRHSLMADCSYTNBEGCKQTEIFGTLTSGTGT ITVPEPLMAVTLQSHAGCRVLAEGRYNTPALANALBHGMAVTVGSATRIEHTCQ WFSHAVKR"

/complement(5094..5099)

/gene="STM1129"

/note="putative RBS for STM1129; ReguonDB:STMS1H001401"

5599..6759

/gene="STM1130"

5599..6759

/gene="STM1130"

/note="similar to E. coli orf, hypothetical protein (AAC77266.1); Blastp hit to AAC77266.1 (404 aa), 41%

identity in aa 40 - 403"

/codon_start=1

/transl_table=1

/product="putative inner membrane protein"

/protein_id="AAL20061.1"

/db_xref="GI:16419646"

/translation="MGMQKPKKMTFLMALCSVAITTSQYATLTDIPPLKNGTG AIDNNGVIVGLGAGTSWKIDLKQKHMERIKSPGGAQREOSVSPFLDELTVRG GVGKNSPQLQVYSDVYKYSPVNTQKQWDTJSPVGLTGTGVKMETVNLITGVNG EHIIDKYFIDIAAADADESEKVKIYVYFNKPADYFPNKIVFIYNKAKENTWAKGEL PACTAGSSSVGMNNPFLMLNGELKPGRTDVIIRAMWDKLTFLKNSQLSPSPGEO QOEGLAGPFGSYSHLVGGGANPFAKQNTYNGKYSHEGINKKRDVYGLINCH MOWYKAKQPLGIVSVSYGDEVLTIGENAKGPVSVTSFTKRDGNLTK"

6793..7497

/gene="STM1131"

6793..6798

/gene="STM1131"

/note="putative RBS for STM1131; ReguonDB:STMS1H001402"

gene /note="PS00142 Neutral zinc metalloproteinases, zinc-binding region signature"
3001..4011
/gene="STY1079"
/note="synonym: PYRD"
3001..4011
/gene="STY1079"
/note="Orthologue of E. coli PYRD (PYRD_ECOLI); Faeta hit to PYRD_ECOLI (336 aa), 95% identity in 336 aa overlap"
/codon_start=1
/transl_table=1
/product="dihydroorotate dehydrogenase"
/protein_id="CAD08184.1"
/db_xref="GI:16502233"
/db_xref="GOA:Q82759"
/db_xref="SWISS-PROT:Q82759"
/translation="MYPPVRKALFQLDPERAHEFTFOQLRITGTPLLEALVROKVP
KPVCMGLTFKRLPLGLAGLDKDECDALGAMGFGSLIGTVPRPOPNDKPLFR
LVDAEGLINRMGFNNLGVNDLVENYKKAHPDGLIGINIGKNDTPVENGRDYLIME
KVAVAGYIAINISSPNTPLGLTQYGDALDPLTALKNKNDLOVTHKKVPAVAKI
APDCEBELIOVADSLAHNIDGVATTTTDDRSIVQMKCCQGTGSGRPLOKST
EITRLSLKGLPLIGVGIDSVIAREKIALAGATLVQISGTFKGPLKEIYV
HI"
3136..3990
/gene="STY1079"
/note="Pfam match to entry PF01180 DHodhase, Dihydroorotate dehydrogenase, score 563.50, E-value 1.4e-165"
3238..3297
/gene="STY1079"
/note="PS00911 Dihydroorotate dehydrogenase signature 1"
3874..3936
/gene="STY1079"
/note="PS00912 Dihydroorotate dehydrogenase signature 2"
4174..4719
/gene="STY1080"
4174..4719
/gene="STY1080"
/note="Orthologue of E. coli YCBM_ECOLI; Faeta hit to YCBM_ECOLI (192 aa), 90% identity in 180 aa overlap. Note discrepancy in position of translational start site."
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08185.1"
/db_xref="GI:16502234"
/db_xref="SPTRMBL:Q82758"
/translation="MMRIKPDNDNRWYDEEDHDMMLDLANGMLFRSFRSKMLTPDA
FCTGFCVDNALYFSFEKCRDELTKGKRAELVNLALVIRYLKQMPKSMHFAVA
GEMWTPGDAASVLSPTAEQVNLVVERBENALCLLAPGVVIAGRTQLDGLAIK
IMDRILKRVCHSFSLEQAV"
complement(4716..5825)
/gene="STY1083"
complement(4716..5825)
/gene="STY1081"
/note="Orthologue of E. coli P75863; Faeta hit to P75863 (369 aa), 86% identity in 367 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08186.1"
/db_xref="GI:16502235"
/db_xref="GOA:Q82757"
/db_xref="SPTRMBL:Q82757"
/translation="MATLSRLFIHPVSMRGIGLTHALADISGLAFRIEMVTSDEGT
FITAROPQWRFTFSPPLDGLHLTAPGSSALVRFIDFTQDDPTEWGNHFAFARA
PAINOMLSGFSDVLCVGPQPLTRKVKHNAVPLGADGYLLTNESLSDLOO
RCPAGVOMEQFRPLVVGVAWAEDSKVIRIGDVPDVKKPSRCIFTTYSBEKQ
KHSGEPALTIQAFRTADNDGVDPCGNLIRNSGVRVSGVEVILATAPKAYGATT
LDSEVTPKRPAGSVTTIDMGQITFCVKNQVLLLEQNGQIRITPISCAIGCGCRIR
LLEGEVPLKKSADMGDTILUSCVPKTAIRLEN"
complement(4737..4955)
/gene="STY1081"

gene /note="Pfam match to entry PF00111 fer2, 2Fe-2S iron-sulfur cluster binding domains, score 55.90, E-value 8.5e-13"
5924..8032
/gene="STY1082"
5924..8032
/gene="STY1082"
/note="Similar to Haemophilus influenzae hypothetical protein h1016 SW:YCBY_HAEIN (P44524; P43945) Faeta scores: E(): 0, 60.3% id in 710 aa
Orthologue of E. coli YCBY_ECOLI; Faeta hit to YCBY_ECOLI (702 aa), 92% identity in 702 aa overlap"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="CAD08187.1"
/db_xref="GI:16502236"
/db_xref="GOA:Q82756"
/db_xref="SPTRMBL:Q82756"
/translation="MNSLPASTARGLSEELKIEKAGVCCOVQGVHFGQTRLI
YQSLMRSRLSRILIPMGCKVTSDDLTFGVQXINTEIFNPATRAVHPSGLNDTI
RNSQYGMKVKDAIVDAFTKRLPRPVNDESPDLRINVMNKETASIALDSDGLH
LRGVRDTGLAPIKETLAAIIVMSGWPGLPLDPMCGSGTLLIEAAMWATDRAPIGL
HRGHMFGSMAOHDEAIWOSVKAEQFARKGLAEYSHFYGSDDARIVSRASNAR
RAGIGELITREVDVAQLSNPLPKPGVGTINPVGEIRLDDSPALIAHSLGRTMK
NORGKMLSPASPDLSIQLEADQFPAKNGPLDVCQKNTIATTTADSKPATYA
EDYANRRLKRLKLEKVARQEGIECVLYADLPEYVAVADRYDMVIOEYAPKTV
DAQKARQRLDIIATLSVGLIPENKLVLTREKQKNOYQKXSEGEFLEVEYVA
RLWNLTDYDLTGFLDRIARMLGMSKGLNLFNFSYSGASVAGAGCATTT
VMSRTYLEMAERMLRLNGSGRAHRLIADCGMLAEANQGLFIIDPPTFNSNR
MESEPDYQRHVALMKDLKRLKRLKRGITIMSNKRGFRMLDGLAEIGLTAQETTKT
LSPDFARNROIHNCMLIRAA"
5927..7060
/gene="STY1082"
/note="Pfam match to entry PF01170 UPF0020, Uncharacterized protein family UPF0020, score 781.30, E-value 3.7e-231"
6506..6541
/gene="STY1082"
/note="PS01261 Uncharacterized protein family UPF0020 signature"
8045..9952
/gene="STY1083"
8045..9952
/gene="STY1083"
/note="Faeta hit to YBIT_ECOLI (530 aa), 32% identity in 526 aa overlap
Faeta hit to YJUK_ECOLI (554 aa), 34% identity in 524 aa overlap

Query Match 88.4%; Score 403.2; DB 1; Length 254050;
Best Local Similarity 92.8%; Pred. No. 5,7e-105;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTTCGACGCTCGGCT 60
Db 88906 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTTCGACGCTCGGCT 88965

QY 61 GCGCTCGTTCACAAATGGGCGGCGGCGTAAATCAACGCGGCAATAGTTCGGC 120
Db 88966 GCGCTCGTTCACAAATGGGCGGCGGCGTAAATCAACGCGGCAATAGTTCGGC 89025

QY 121 CCGGACTCAACGTTGAGCATTTATCAGTAGCGTTCCGCTAACGCTGGCTTCTGCA 180
Db 89026 CCGGACTCAACGTTGAGCATTTATCAGTAGCGTTCCGCTAACGCTGGCTTCTGCA 89085

QY 181 AGCGATCCCGTAAATATGATACGCTGTTAACCGGTGTGTTACCACTGAAGATGAT 240
Db 89086 AGCGATCCCGTAAATATGATACGATTAACCAAGCGGTTATGTTACCGGCGAT 89145

QY 241 GCAGCGCAGGAGTGGGATTAATAGTACTATTGAACCTGACTGAGATGGTTTCAGAAATAT 300
Db 89146 GTAGCGCAGGAGTGGGATTAACGATTAATGAACCTGACTGAGATGGTTTCAGAAATAT 89205

[illegible]

gene
complement(5597..6295)
/gene="flgH"
/locus_tag="c1740"
complement(5597..6295)
/gene="flgH"
/locus_tag="c1740"
/note="corresponds to STY1219 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQRYALHAYPMALMATLTCGAMIPAKPLVQCATTTQAPLPGV
PVANSITQSAQPIVYGTQPLFEDRRPNIGTTLTVQENVSASSANASRDGKT
SFGDTVPRIYLOGLFNSRADMEASGSPFKGKASNTFSGLTVTVQVLANGN
LHVVEKQIAINQCTEFIRFSGVNPRTISGNSVSTQVADARIEVYNGYINBAON
MGMLQRFNLSPM"
complement(6350..7132)
/gene="flgG"
/locus_tag="c1741"
complement(6350..7132)
/gene="flgG"
/locus_tag="c1741"
/note="distal rod protein: corresponds to STY1218 from
Accession AU513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISLSLWAKTGELDAQNTMDVIANNLAVSTNGSKRQRAVEDL
LYQIRQGAQSEDTLPSSQIGTGVPAATERLHSGMLSTNNKSDVAIKQGF
FOVMLPDGTSAYTRDGSFQVQNGQLVAGQVQVPAITIPANALSTIGRDGVSVT
QOQGAAPVQVQLMTTFMNDTGLSEIGNENLITQSSGAPNESTPGINGAGLLYQGY
VETSNVAEEIVMMIOVRAVEINSKAVSTTDQMLQKLTQL"
complement(7146..7901)
/gene="flgF"
/locus_tag="c1742"
complement(7146..7901)
/gene="flgF"
/locus_tag="c1742"
/note="corresponds to STY1217 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHALITRMGASQTLNQAVTASNLANASTPGRALNALRAV
FVDSLATRTLVASTPGADMTQGLDYSRPLDVALQOQGMVLVQADAEQTRN
GNIOVPTQGLTIOGHPIVIGGGPITVEGSEITADAGTISALNPGDPPTVAIPVR
LKLVAEGNEVQRSDGLFRLTAERGAERGAIVLADPSIRIMSGVLBSNPKVPEAMT
DMANARFEMQMKVITSVDNEGRANQLLSMS"
complement(7922..9133)
/gene="flgE"
/locus_tag="c1743"
complement(7922..9133)
/gene="flgE"
/locus_tag="c1743"
/note="corresponds to STY1216 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSPSOAVSGLNAAATNLDTGNNTANSATYGFKSGTASPADMA
GSKVGLGVKAVGITQDFDGTNTNGRLDVAISONGFPRLVDSNGSVFYSRNQDFKL

gene
complement(5597..6295)
/gene="flgH"
/locus_tag="c1740"
complement(5597..6295)
/gene="flgH"
/locus_tag="c1740"
/note="corresponds to STY1219 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar L-ring protein precursor"
/protein_id="AA069364.1"
/db_xref="GI:29137803"
/translation="MQRYALHAYPMALMATLTCGAMIPAKPLVQCATTTQAPLPGV
PVANSITQSAQPIVYGTQPLFEDRRPNIGTTLTVQENVSASSANASRDGKT
SFGDTVPRIYLOGLFNSRADMEASGSPFKGKASNTFSGLTVTVQVLANGN
LHVVEKQIAINQCTEFIRFSGVNPRTISGNSVSTQVADARIEVYNGYINBAON
MGMLQRFNLSPM"
complement(6350..7132)
/gene="flgG"
/locus_tag="c1741"
complement(6350..7132)
/gene="flgG"
/locus_tag="c1741"
/note="distal rod protein: corresponds to STY1218 from
Accession AU513382: Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar basal-body rod protein flgG"
/protein_id="AA069365.1"
/db_xref="GI:29137804"
/translation="MISLSLWAKTGELDAQNTMDVIANNLAVSTNGSKRQRAVEDL
LYQIRQGAQSEDTLPSSQIGTGVPAATERLHSGMLSTNNKSDVAIKQGF
FOVMLPDGTSAYTRDGSFQVQNGQLVAGQVQVPAITIPANALSTIGRDGVSVT
QOQGAAPVQVQLMTTFMNDTGLSEIGNENLITQSSGAPNESTPGINGAGLLYQGY
VETSNVAEEIVMMIOVRAVEINSKAVSTTDQMLQKLTQL"
complement(7146..7901)
/gene="flgF"
/locus_tag="c1742"
complement(7146..7901)
/gene="flgF"
/locus_tag="c1742"
/note="corresponds to STY1217 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="putative flagellar basal-body rod protein flgF"
/protein_id="AA069366.1"
/db_xref="GI:29137805"
/translation="MDHALITRMGASQTLNQAVTASNLANASTPGRALNALRAV
FVDSLATRTLVASTPGADMTQGLDYSRPLDVALQOQGMVLVQADAEQTRN
GNIOVPTQGLTIOGHPIVIGGGPITVEGSEITADAGTISALNPGDPPTVAIPVR
LKLVAEGNEVQRSDGLFRLTAERGAERGAIVLADPSIRIMSGVLBSNPKVPEAMT
DMANARFEMQMKVITSVDNEGRANQLLSMS"
complement(7922..9133)
/gene="flgE"
/locus_tag="c1743"
complement(7922..9133)
/gene="flgE"
/locus_tag="c1743"
/note="corresponds to STY1216 from Accession AU513382:
Salmonella typhi CT18"
/codon_start=1
/transl_table=11
/product="flagellar hook protein flgE"
/protein_id="AA069367.1"
/db_xref="GI:29137806"
/translation="MSPSOAVSGLNAAATNLDTGNNTANSATYGFKSGTASPADMA
GSKVGLGVKAVGITQDFDGTNTNGRLDVAISONGFPRLVDSNGSVFYSRNQDFKL

Query Match 88.4%; Score 403.2; DB 1; Length 301983;
Best Local Similarity 92.8%; Pred. No. 5.8e-105;

Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTTCGCGAGTCTGGCT 60
DB 37310 ATGAACTTTTAAAGTGGCAGCATTTGGCAGCAATCGTAGTTTTCGCGAGTCTGGCT 37251
QY 61 GGGCGTTCACAAATGGGGGGGGGGGTAAATCAATACGGCGGCGGCAATAGTTCCGGC 120
DB 37250 GGGCGTTCACAAATGGGGGGGGGGGTAAATCAATACGGCGGCGGCAATAGTTCCGGC 37191
QY 121 CCGACATCAACGTTAGCATTTATAGTACCGTTCCGTTACGTTCCGTTCCGTTCCGTT 180
DB 37190 CCGACATCAACGTTAGCATTTATAGTACCGTTCCGTTACGTTCCGTTCCGTTCCGTT 37131
QY 181 AGCGATGCGGTTAAATATGATCAGCTGGTTACCGTGTGTTTACCATGAATGCGCAT 240
DB 37130 AGCGATGCGGTTAAATATGATCAGCTGGTTACCGTGTGTTTACCATGAATGCGCAT 37071
QY 241 GCAGGCCAGGGTGGGATTAATAGTACTATTGAATGAGCTGCAATGTTTCAGAAATAT 300
DB 37070 GTAGCCAGGGTGGGATTAATAGTACTATTGAATGAGCTGCAATGTTTCAGAAATAT 37011
QY 301 GCCACATCGACAGTGGAACTGTAATATCTGTCGCGCATATACGCGCGT 360
DB 37010 GCCACATCGACAGTGGAACTGTAATATCTGTCGCGCATATACGCGCGT 36951
QY 361 AATAACGCGCGTGTGTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGTT 420
DB 36950 AATAACGCGCGTGTGTTATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGTT 36891
QY 421 GGTTCGCAACAGCGCAGCGCTAACAGTATTA 456
DB 36890 GGTTCGCAACAGCGCAGCGCTAACAGTATTA 36855

RESULT 7
STAGFA
LOCUS STAGFA 1048 bp DNA linear BCT 26-JAN-1998
DEFINITION Salmonella typhimurium agfB and agfA genes.
ACCESSION AU000514
VERSION AU000514.1 GI:2275119
KEYWORDS agfA gene; agfB gene.
SOURCE
ORGANISM Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE
AUTHORS Sukupolvi,S., Lorenz,R.G., Gordon,J.I., Bian,Z., Pfeiffer,J.D.,
Normark,S.J., and Rhen,M.
TITLE Expression of thin aggregative fimbriae promotes interaction of
Salmonella typhimurium SR-11 with mouse small intestinal epithelial
cells
JOURNAL Infect. Immun. 65 (12), 5320-5325 (1997)
MEDLINE 98053981
PUBMED 9393832
REFERENCE 2 (bases 1 to 1048)
AUTHORS Sukupolvi,S.S.
TITLE Direct Submission
JOURNAL Submitted (14-JUL-1997) Sukupolvi S.S., Medical Biochemistry,
University of Turku, Kiinamyllynkatu, 20520, FINLAND
FEATURES
SOURCE
1..1048
/organism="Salmonella typhimurium"
/mol_type="genomic DNA"
/db_xref="taxon:602"
14..553
/gene="agfB"
14..553
/gene="agfB"
/codon_start=1
/transl_table=11
/product="AgfB protein"
/protein_id="CAA04150.1"

```

/db_xref="GI:2275120"
/db_xref="SPTREMBL:033801"
/translation="MKNLFLPMLTJLGEPIATATNYDLARSEYFAVNELSKSFNO
AAITGVGTDSNARVHRHSGSLISVIOERGGNNRAVDAGNYNPAFYEIOTGNADA
SISQAVGNSSAIIOKSGNKNKITGYQTKFVVVQKSHALQANITQYQTO
KTAVVQKSHAIRVTR"
593..1048
/gene="agfa"
593..1048
/cds
/gene="agfa"
/codon_start=1
/translation="Agfa protein"
/protein_id="CA04151.1"
/db_xref="GI:2275121"
/db_xref="SPTREMBL:033802"
/translation="MKLKVAFPAIVSGSAVAVVPQWGGGNNHNGSGSDST
LSIYQGSANAALLOSADKSEITITQSYGAGVADGGAGNISTELTONFRNNAT
IDWNNANSIDITVGYGGNNMALVNOTASDSVMVQVGFNNAPANGYN"
ORIGIN

Query Match      87.1%; Score 397.4; DB 1; Length 1048;
Best Local Similarity 92.1%; Pred. No. 1.9e-103;
Matches 419; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTGTGTCGAGTCTCGGCT 60
DB 593 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTGTGTCGAGTCTCGGCT 652
QY 61 GGGCTGTTTCAATGGGGGGGGGGGGTATTCATACGGCGGCAATGTTCCGGC 120
DB 653 GGGCTGTTTCAATGGGGGGGGGGGGTATTCATACGGCGGCAATGTTCCGGC 712
QY 121 CCGGACTCAACGTTTGAATATCAGTACGGTCCGTAACGTCGCGCTGCTCGAA 180
DB 713 CCGGACTCAACGTTTGAATATCAGTACGGTCCGTAACGTCGCGCTGCTCGAA 772
QY 181 AGCGATGCCCGTAATATGATCAGTCTGTTACCGCTGTTTACCATGAATGCAAT 240
DB 773 AGCGATGCCCGTAATATGATCAGTCTGTTACCGCTGTTTACCATGAATGCAAT 832
QY 241 GCAGGCGCAGGTGCGGATATGATCTATTGAACTGACTGAGATGGTTTCAGAAAT 300
DB 833 GTAGGCGCAGGTGCGGATATGATCTATTGAACTGACTGAGATGGTTTCAGAAAT 892
QY 301 GCCACCATGACGATGGAAGCTTAAACTCCATATTAATGTCGCGCAATACGGCGT 360
DB 893 GCCACCATGACGATGGAAGCTTAAACTCCATATTAATGTCGCGCAATACGGCGT 952
QY 361 AATAAGCGCGGCTGTTATCAAGCGCATCTGATCCAGCGTAATGTCGCTCAGGT 420
DB 953 AATAAGCGCGGCTGTTATCAAGCGCATCTGATCCAGCGTAATGTCGCTCAGGT 1012
QY 421 GGTTCGCAACAGCCGCGCTTACCATGATTA 455
DB 1013 GGTTCGCAACAGCCGCGCTTACCATGATTA 1047

RESULT 8
LOCUS 144908 361 bp DNA linear PAT 07-OCT-1997
DEFINITION Sequence 56 from patent US 5635617.
ACCESSION 144908
VERSION 144908.1 GI:2469621
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
REFERENCE
1 (bases 1 to 361)
AUTHORS Doran,J.L., Kay,W.M., Collinson,S.Karen. and Cloughier,S.C.
TITLES Methods and compositions comprising the agfa gene for detection of
JOURNAL Patent: US 5635617-A 56 03-JUN-1997;

```

```

FEATURES
source
location/Qualifiers
1..361
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN

Query Match      63.3%; Score 288.6; DB 6; Length 361;
Best Local Similarity 91.3%; Pred. No. 4.5e-72;
Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTGCTTCCAAATGGGGGGGGGGGGTATCATACGGCGGCAATGTTCCGGCCG 123
DB 1 GTGCTTCCAAATGGGGGGGGGGGGTATCATACGGCGGCAATGTTCCGGCCG 60
QY 124 GACTCAACGTTGAGCAATTAATCACTACGTTCCGCTTAAAGCTTCTCTGCAAGC 183
DB 61 GACTCAACGTTGAGCAATTAATCACTACGTTCCGCTTAAAGCTTCTCTGCAAGC 120
QY 184 GATGCCCGTAATATGATCAGTCTGTTACCGCTGTTTACCATGAATGSCACATCA 243
DB 121 GATGCCCGTAATATGATCAGTCTGTTACCGCTGTTTACCATGAATGSCACATCA 180
QY 244 GGGCAGGGTCCGATATGATCTATTGAACTGACTGAGATGTTTCAAAATATGCC 303
DB 181 GGGCAGGGTCCGATATGATCTATTGAACTGACTGAGATGTTTCAAAATATGCC 240
QY 304 ACCATGACGATGGAAGCTTAAACTCCATATTAATGTCGCGCAATACGGCGTAT 363
DB 241 ACCATGACGATGGAAGCTTAAACTCCATATTAATGTCGCGCAATACGGCGTAT 300
QY 364 AACGCGCGCTGTTATCAACCGCATCTGATTC 398
DB 301 AACGCGCGCTGTTATCAACCGCATCTGATTC 335

RESULT 9
LOCUS CSP515700 2889 bp DNA linear BCT 24-JUN-2003
DEFINITION Citrobacter sp. Fec2 csb gene, csb gene and csd gene.
ACCESSION AJ515700
VERSION AJ515700.1 GI:31790491
KEYWORDS csb gene; csb gene; csd gene; curlin-csb protein; nucleation
component of curlin monomers; regulatory protein.
SOURCE
ORGANISM
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Citrobacter.
1
Zogaj,X., Bokranz,W., Nimetz,M. and Romling,U.
Production of Cellulose and Curli Fimbriae by Members of the Family
Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
Infect. Immun. 72 (7), 4151-4158 (2003)
JOURNAL
2 (bases 1 to 2889)
AUTHORS Romling,U.
TITLES Direct Submission
SUBMITTED (11-NOV-2002) Romling U., Microbiology and Tumorbiology
Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
JOURNAL
location/Qualifiers
1..2889
/organism="Citrobacter sp. Fec2"
/mol_type="genomic DNA"
/isolate="Fec2"
/isolation_source="faeces"
/specific_host="Homo sapiens"
/db_xref="taxon:213763"
/country="Germany"
/complement(212..862)
/gene="csbD"
/complement(212..862)
/gene="csdD"
/function="necessary for transcription of the csbA
operon"
/codon_start=1

```

```

/transl_table=11
/product="putative regulatory protein"
/protein_id="CAD56670.1"
/db_xref="GI:31790492"
/translation="MFNEFHSIHGHTLLITKPSIQATALLQHKQSIALSGKLNINQ
RSLDDISSCIVLDMMEDKLIHYMODNISRKNNIKTLILNTPDYPRDIENWP
HINQFVYADDEERYNGLQILRECFYSGOKLASLITTHSGNRYNSSEALITRE
KEILNKIRIGASNIENIARSLFISENTVTHLYNLFKIAVNRQAVSWANDNLR"
1623..2078
/gene="csgb"
1623..2078
/gene="csgb"
/function="nucleator"
/codon_start=1
/transl_table=11
/product="nucleation component of curlin monomers"
/protein_id="CAD56671.1"
/db_xref="GI:31790493"
/translation="MKNKLLFMMLTVLGAPGIASATSYDLAHSEYNPAVNELSKSPN
OAILIGVGTNNSAKMRQESKLLSYVSGESGNARAKADOGANFAVTAIQSGHSNDA
SISOSNGTMTMIIOKSGSNKANTQYGTQTAIVVQKQSGMAIRVQR"
2119..2571
/gene="csga"
2119..2571
/gene="csga"
/function="curli subunit"
/codon_start=1
/transl_table=11
/product="curlin-csgA protein"
/protein_id="CAD56672.1"
/db_xref="GI:31790494"
/translation="MKLLQVAPFAIVVSGSALAGSVPMQGGGGGGSGSGSESTL
SIYSGVNNALALQSDPKRSPTTIHQGFENGADVGSGSNTSIDLQNFKNNTTI
DOMNGKSDITVSOTIGHNALVNOTASDSSVLVHQVGFNNATNQT"

ORIGIN
Query Match      61.4% Score 280; DB 1; Length 2889;
Best Local Similarity 77.4%; Pred. No. 1,66-69;
Matches 353; Conservative 0; Mismatches 100; Indels 3; Gaps 1;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACCAATCGTAGTTTCTGGCAGTCTTGCT 60
Db 2119 ATGAAACTTTTAAAGTGGACGATTCGACCAATCGTAGTTTCTGGCAGTCTTGCT 2178
QY 61 GGCCTCGTTCCACATAGGGGGGGCGCGGTATCTAATACGGCGGCGCAATAGTTCGGC 120
Db 2179 GGTTCGTGTTCCGCAATGGGGGGCGCGCGG---TGGCGGGCGGGGAGCAAGCTCCGGC 2235
QY 121 CCGGACTCAAGCTTATGAGCATTTATCAGTACGAGTTCCGCTAACGGTGGCTGCTGCA 180
Db 2236 CCGGACTCAAGCTTATGAGCATTTATCAGTACGAGTTCCGCTAACGGTGGCTGCA 2295
QY 2236 CCGGACTCAAGCTTATGAGCATTTATCAGTACGAGTTCCGCTAACGGTGGCTGCA 2295
Db 181 AGCGATGCCCTTAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAATGGCAAT 240
QY 2296 AGCGATGCCCTTAATATGATCAGCTGGTTACCCGTGTTGTTACCCATGAATGGCAAT 2355
Db 241 GCGAGCCCGGGTGGCGATTAATAGTACTTGAACCTGATCGATGCTTTCAGAAATTAAT 300
QY 241 GCGAGCCCGGGTGGCGATTAATAGTACTTGAACCTGATCGATGCTTTCAGAAATTAAT 300
Db 2356 GTGGCCCGAGGCTCAGATTAACGACCATCATCTGACTCAAAACGGGCTTCAAAAACAC 2415
QY 301 GCCACATCGACGAGTGAACGCTAAAACCTCCGATATTAATGTCGGCCAAATACGGCGGT 360
Db 2416 GCCACATCGATCACTGAGAACGGCAAAATTCGACATTAATCTGAGCCAGTAATGTGGA 2475
QY 361 AATAACGGCGGCTGTTAATCAACCGCATCTGATTCACGCGTAATGTCGTCAGGTT 420
Db 2476 CATTAACGGCGGCTGTTAATCAACCGCATCTGATTCACGCGTTCTGTGCATCAGGTT 2535
QY 421 GGTTTGGCAACAACGCGCAAGCTTAACCAATTA 456
Db 2536 GGTTTGGCAACAACGCGCAAGCTTAACCAATTA 2571

```

```

RESULT 10
ECCSGABDG      4680 bp      DNA      linear      ECT 07-JUL-2002
LOCUS          E.coli csgG, csgF, csgE, csgD, csgB, csgA, and orfC genes.
DEFINITION     X90754
ACCESSION      X90754.1
VERSION        X90754.1 GI:1147558
KEYWORDS       csgA gene; csgB gene; csgD gene; csgE gene; csgF gene; csgG gene;
               orfC gene;
SOURCE          Escherichia coli
ORGANISM       Escherichia coli
REFERENCE      1
AUTHORS        Hammar, M., Arngvist, A., Bian, Z., Olsen, A. and Normark, S.
TITLE          Expression of two csg operons is required for production of
               fibronectin- and congo red-binding curli polymers in Escherichia
               coli K-12
JOURNAL        Mol. Microbiol. 18 (4), 661-670 (1995)
MEDLINE        96414468
PUBMED         8817489
REFERENCE      2 (bases 1 to 4680)
AUTHORS        Hammar, M.
TITLE          Direct Submission
JOURNAL        Submitted (11-AUG-1995) M. Hammar, Karolinska Institutet,
               Microbiology and Tumorbiology Center, Box 280, S-171 77 Stockholm,
               SWEDEN

FEATURES
source
    location/Qualifiers
        1..4680
        /organism="Escherichia coli"
        /mol_type="genomic DNA"
        /strain="K12"
        /db_xref="taxon:562"
        /map="23.15 minutes"
        1..3140
        /organism="Escherichia coli"
        /mol_type="genomic DNA"
        /strain="MC4100"
        /db_xref="taxon:562"
        3141..4680
        /organism="Escherichia coli"
        /mol_type="genomic DNA"
        /strain="N3110"
        /db_xref="taxon:562"
        <1..34680
        /insertion_seq="IS2 (partial)"
        complement(133..966)
        /gene="csgG"
        complement(133..966)
        /gene="csgG"
        /codon_start=1
        /transl_table=11
        /product="assembly /transport component in curli
        production"
        /protein_id="CA62277.1"
        /db_xref="GI:1147559"
        /db_xref="COA:PS2103"
        /db_xref="SWISS-PROT:P52103"
        /translation="MORFLFLVAVMLSGCLTAPPKAAAPTLMPPAQSRYKDLTHLPA
        PTGKIFVSVNIDDETQCFKPYPASNSFTAVPOSATMLVTALKDSRMTPLERQGLQ
        NLNMRKIRAAQNGTVAINNRIPLQSLTAANIMVSGSLIGVESNVSGGVGARVYG
        IGDVTOLODOIIVNLRVNVNVTGELISSVNTSKTILSYEVOAGVPRFIYORLLEGB
        VGYTSNRPVNLCLMSALETGVIFLINDIDRGMLDQNKAKRONDIIVKTRHNSVPE
        S"
        complement(993..1409)
        /gene="csgF"
        complement(993..1409)
        /gene="csgF"
        /codon_start=1
        /transl_table=11
        /product="assembly /transport component in curli
        production"
        /protein_id="CA62278.1"
        /db_xref="GI:1147560"

```


University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-262-7459

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 3 (baees 1 to 10346)
 Blattner, F.R.
 Direct Submission
 Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
 University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 Email: ecolig@genetics.wisc.edu Phone: 608-262-2534 Fax: 608-262-7459

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 4 (baees 1 to 10346)
 Plunkett, G. III.
 Direct Submission
 Submitted (13-OCT-1998) Laboratory of Genetics, University of
 Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

COMMENT
 This sequence was determined by the E. coli Genome Project at the
 University of Wisconsin-Madison (Frederick R. Blattner, director).
 Supported by NIH grants HG00301 and HG01428 (from the Human Genome
 Project and NCHGR). The entire sequence was independently
 determined from E. coli K12 strain MG1655. Predicted open reading
 frames were determined using Genemark software, kindly supplied by
 Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
 30332 (e-mail: markborov@gatech.edu). Open reading frames that
 have been correlated with genetic loci are being annotated with CG
 Site Nos., unique ID nos. for the genes in the E. coli Genetic
 Stock Center (CGSC) database at Yale University, kindly supplied by
 Mary Berlyn. A public version of the database is accessible
 (<http://cgsc.biology.yale.edu>). Annotation of the genome is an
 ongoing task whose goal is to make the genome sequence more useful
 by correlating it with other data. Comments to the authors are
 appreciated. Updated information will be available at the E. coli
 Genome Project's World Wide Web site
 (<http://www.genetics.wisc.edu>): *** The E. coli K12 sequence and
 its annotations are periodically updated; this is version M54. No
 sequence changes. Annotation updates: updated gene identifications
 and products; all new functional assignments courtesy of Monica
 Riley; added promoters, protein binding sites, and repeated
 sequences described in reference 1. The unique numeric identifiers
 beginning with a lowercase 'b' assigned to each gene (protein- or
 RNA-encoding) are now designated as gene synonyms instead of
 labels. This should allow them to be searched for in Entrez as gene
 names.

FEATURES
 source
 1..10346
 /organism="Escherichia coli K12"
 /mol_type="genomic DNA"
 /strain="K12"
 /sub_strain="MG1655"
 /db_xref="taxon:83333"
 complement(1..48)
 /note="IS3"

repeat_region
 1..5
 /note="Factor Sigma70; predicted +1 start at 1094689"

promoter
 5..32
 /note="Factor Sigma70; predicted +1 start at 1094716"

gene
 69..392
 /gene="b1028"
 69..392
 /gene="b1028"
 /function="orf; Unknown"
 /note="0107; This 107 aa ORF is 27 pct identical (0 gaps)
 to 58 residues of an approx. 188 aa protein PAS2_CANAL
 SM: P43098"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74113.1"
 /db_xref="GI:1787266"
 /translation="MEDGVQAMRDYLAGLDIASPEHOYLVNVTAKSEVAPSIKENIS
 LHLTHYKWTESLDTFLLNPTVAFLEISNRPYIGNMLNDPAGVDQQRVIMHCRAFSD
 AKVPK"
 275..299
 /gene="b1028"

promoter

gene
 389..1375
 /gene="ycdu"
 /note="synonym: b1029"

CDS
 389..1375
 /gene="ycdu"
 /function="orf; Unknown"
 /note="0328; This 328 aa ORF is 19 pct identical (4 gaps)
 to 191 residues of an approx. 288 aa protein Y320_MYCE
 SM: P47562"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74114.1"
 /db_xref="GI:1787267"

protein_bind
 1253..1280
 /note="central position to predicted promoter: -104"
 /bound_moiety="Deor predicted site"

protein_bind
 1253..1280
 /gene="ycdu"
 /note="No predicted promoter"
 /bound_moiety="Deor predicted site"

promoter
 1337..1365
 /gene="ycdu"
 /note="Factor Sigma70; predicted +1 start at 1096049"

gene
 1494..1745
 /gene="b1030"

CDS
 1494..1745
 /gene="b1030"
 /function="orf; Unknown"
 /note="083; This 83 aa ORF is 25 pct identical (7 gaps)
 to 81 residues of an approx. 616 aa protein ALBU_Rab1t SW:
 P49065"
 /codon_start=1
 /transl_table=1
 /product="orf, hypothetical protein"
 /protein_id="AAC74115.1"
 /db_xref="GI:1787268"

gene
 1513..1926
 /translation="MGGTRCVOYTKKRPVLSYELFPKYGEGGIDLSPPCGPAHC
 VHGISNMLSPVVDPRSGLLIPVCAIHEKKARTEFRALL"
 1513..1926
 /gene="ycdv"
 /note="synonym: b1031"

CDS
 1513..1926
 /gene="ycdv"
 /function="putative structure; Not classified"
 /note="0137; This 137 aa ORF is 27 pct identical (1 gap)
 to 47 residues of an approx. 80 aa protein RS37_YEAST SW:
 P05759"

gene
 205759
 /codon_start=1
 /transl_table=1
 /product="putative ribosomal protein"
 /protein_id="AAC74116.1"
 /db_xref="GI:1787269"

gene
 205759
 /translation="MCNIRKKSYPFRSSLSNMAVNGGLTRFAPRAGSPITATVCP
 GROLSTPGRFSRSPRCVQTKKKPVLSYELFPKYGEGGIDLSRSPCGPAHCVHGL
 SNMLSPVPEPRSGLLIPVCAIYEKKARTEFRALL"
 complement(2111..2198)
 /gene="serx"
 /note="synonym: b1032"

gene
 complement(2111..2198)
 /gene="serx"
 /product="tRNA-Ser"

tRNA
 /note="anticodon: GGA; CG Site No. 17878"

promoter
 2269..2296
 /function="RNA; tRNA"

Initiation site of ThrA (0 min.).-This clone is from Kohara lambda minisec library."

<1..276
/note="unnamed protein product; ORF_ID:0230#5"
/codon_start=1
/transl_table=1
/protein_id="BAJ35822.1"
/db_xref="GI:4062605"
/translation="DLFPSSSENAISFPFAKPLMLRFLVLIVEMWITLVSVATLVE
WRIPFAVFLCYVCLLYFFPCFLMNDPMACDMIFRMGHFNK"
1152..2090
/note="ORF_ID:0231#1
similar to SwissProt Accession Number P45637"
/codon_start=1
/transl_table=1
/product="Hypothetical 33.0 kD protein in prob-proA
intergenic region."
/protein_id="BAJ35823.1"
/db_xref="GI:4062606"
/translation="WDIIFVHTPTDOWMIEALRKAIPOAVPRAMKSGDMSADYAV
WHPVEMLAGDKAVFALGAGVDSILSKLOHPMLNPSPFLPRLBTGSGEQMERY
AVSQVLMFRFRPDDRIQNSHWPPLPEYHEDFTIGILAGVLSKVAOSLQWRF
PLRCWSRTKSMGPVQSPAGRELSAFLSQCRVLIPLPTPETVGIINQOLEKLPD
GAYLTLARGVHVEDDLAALDSGKGMALDVNREPLPESPFLMQHPVITTPHV
AATRPAAVEVISRTIAQLEKGERVCGVDPRAGY"
2145..2882
/note="unnamed protein product; ORF_ID:0231#2"
/codon_start=1
/transl_table=1
/protein_id="BAJ35824.1"
/db_xref="GI:4062607"
/translation="WYVDLHMHTVASTHAYSTLSDYIAOKKGIKLFATIDGPPDM
EDAPHHMFRIMRIWPRVVDGVLIRGIEANIKVDSIICSGKRFSDLDLIAGFHE
FVPRPHDPAITQAMITATISGNVHITSHPGPKETIDKVAEAAHQAQVALEPNS
SFLSRKSGEDNCREVAARVADAGGWAFLGSDSHAFPTMGFEBCLKILDAVDPPER
ILNVSPRLNLFLESGMAPIAEFADL"
2906..3460
/note="ORF_ID:0231#3
similar to PIR Accession Number F64035"
/codon_start=1
/transl_table=1
/product="Hypothetical protein H11543"
/protein_id="BAJ35825.1"
/db_xref="GI:4062608"
/translation="WNESTICRVIGSYRQPDPLVPLFLIREKLANPPLQ
DELRLQKSCDMQVSADYVLFIDGCAFPYRSANVEATEVAEVAFLSEKGMPL
ADTPADHIGTLAASWLEQSTEDSESLTLPSEYLLPWCGAFLGKVEAHATTPEW
RTMAPLTRDAISAMWDELEBDEE"
3514..4053
/note="unnamed protein product; ORF_ID:0231#4"
/codon_start=1
/transl_table=1
/protein_id="BAJ35826.1"
/db_xref="GI:4062609"
/translation="WCSISFMAAFSAIRKGNLILSLAITTGILSGIWVAVSLGL
SMAGIGCTAFACPGGSLKGLAISATLLSGVWAMWIIYIGSLAPLETILGYVITG
IVAPLMCTQAKOLISFYPGTFICACATPAQGDWKLVPSELALGLITGYMKNSGLM
LAARSAKTAHREQETKNA"
complement (4117..4950)
/gene="cs8g"
complement (4117..4950)
/note="ORF_ID:0231#5
similar to SwissProt Accession Number P52103"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs8g precursor."
/protein_id="BAJ35827.1"
/db_xref="GI:4062610"
/translation="WQRLFLVAVWMLSGCLTAPPEKARPTLMPRAGSYDLTHLPA
PTGKIPVSVNIIODETGFQKPYPASNPATVAPOSATNALVALDLSRPFIDLEQGLQ
NLNBRKTIIRAAOENGVAINNRIPLQSLTAANIMVEGSIIGESNVSGGVGARVFG

IGADTOYOLDIIVANLEFVNVSFGEILLSVTSKTLISYEVOAGVFRIDYORLLEGE
VGYTSNBPVWELCLMSALETGYIFLINDIGRGLDQNKARQNDIIVKRYKMSVPE
S"
complement (4977..5393)
/gene="cs8g"
complement (4977..5393)
/gene="cs8g"
/note="ORF_ID:0231#6
similar to SwissProt Accession Number P52104"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs8g precursor."
/protein_id="BAJ35828.1"
/db_xref="GI:4062611"
/translation="WRVHAVALMLISPLSWAGTWTFQFPNPNFGNPNNGAFLLNS
AQOANSYKDPBYNDPGLIETPSALDNFQALQSOIILGGLLSINITGKRGKRWTDYIV
DIARNDQOLQIANDRKGTGTSITQVSGLIQNNSTDF"
complement (5418..5807)
/gene="cs8g"
complement (5418..5807)
/note="ORF_ID:0231#7
similar to SwissProt Accession Number P52105"
/codon_start=1
/transl_table=1
/product="Curli production assembly/transport component
Cs8g precursor."
/protein_id="BAJ35829.1"
/db_xref="GI:4062612"
/translation="MKRYLRWIVAAEFLPAAGNLHAVEVYRGLTDHTVSSIGDFY
RASPDKESDYTGNTLTINERPSAWGSMWITITVNOVDVIFQFPLPLKRDPEKTVFAL
IQTEBALNRKQINQALSTGDLADEP"
complement (5812..6462)
/gene="cs8g"
complement (5812..6462)
/note="ORF_ID:0231#8
similar to SwissProt Accession Number P52106"
/codon_start=1
/transl_table=1
/product="Probable cs8g operon transcriptional regulatory
protein."
/protein_id="BAJ35830.1"
/db_xref="GI:4062613"
/translation="WNEBVSIGHITLLITKSLQATPALLQHLKQSLATGKLMNQ
RSLDDISGSIILDMMEADKLIHYQDPLSRKNNI KILMLTPEDYPRDIENWP
HINGVFYSMEQERVVNGLOVLRGECYFQKLASVYLITSGNRYNSTESALNLTRES
KEILNKLRIQSGSNNEIARSLFISENTVKTLYNLFKKIAVKNRTQAVSAMANDLNR"
7217..7672
/gene="cs8g"
7217..7672

Query Match 52.3%; Score 238.4; DB 1; Length 15047;
Best Local Similarity 70.2%; Pred. No. 1.7e-57;
Matches 320; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTGCGACAAATCGTAGTTCTGCGAGTCTGCT 60
DB 7713 ATGAACTTTTAAAGTGGCAGATTGCGACAAATCGTAGTTCTGCGAGTCTGCGA 7772
QY 61 GCGCTGCTCCACAAATGCGCGCGCGCGGCTAATACATACGCGCGCGCAATAGTCCGCG 120
DB 7773 GGTGTGTTCTCGTACGTCGCGCGCGCGGCTAATACATACGCGCGCGCAATAGTCCGCG 7832
QY 121 CCGGACTCAAGCTTGAGCATTTATCAGTACGGTTCCTGCTACGCTGCGCTGCTGCA 180
DB 7833 CCAATTTCTGAGCTGAACATTATTCAGTACGGTTCGCGTACCTGCACTTCTGCA 7892
QY 181 AGCAATGCCGCTAAATATGATCAGCTGGTTACCCGTGTTGTTACCAATGAATGGACAT 240
DB 7893 ACTGATGCCGCTAATCTGACTTACTATTAACAGATGCGCGCGGATATGATGTCAGAT 7952

QY 241 GCAGCGCAGGCGGCGGATTAATAGTACTATGTAAGTACTGACGAATGGTTTCAGAAATATAT 300
 DB 7953 GTTGGTCAAGGCTCAGATGACAGCTCAATCATCTGACCCAACTGGCTTCGTAAACAGC 8012
 QY 301 GCCACCATGACGACGCTGAGAGCTTAAACTCCGATTTACTGTCCGCAATACGGGGT 360
 DB 8013 GCTACTCTTGATCACTGAGACGGCAAAATCTTAATGACGGTTAAACAGTTGGGGT 8072
 QY 361 AATACGCGCGCGCTGCTTATATCAACCGCATCTGATTCAGCGTAAATGGTCCGTCAGATT 420
 DB 8073 GGCACAGGCTGCGAGTTCAGACCTGACCTTAATCTCCGTCACAGTCACTGAGGTT 8132
 QY 421 GGTTTGGCAACAACGCGCAGCTTACAGATTTAA 456
 DB 8133 GGGTTTGGTACAAACGCGACCGCTCATCAGTACTTA 8168

RESULT 13

CFS15701 2920 bp DNA linear BCT 24-JUN-2003
 LOCUS Citrobacter freundii csb gene, csb gene and csd gene.
 DEFINITION A515701
 ACCESSION A515701.1 GI:31790495
 VERSION csb gene, csd gene, curlin-csb protein; nucleation
 KEYWORDS component of curlin monomers; regulatory protein.
 SOURCE Citrobacter freundii
 ORGANISM Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Citrobacter.
 1 Zogaj, X., Bokranz, W., Nitz, M. and Romling, U.
 REFERENCE Production of Cellulose and Curli Fimbriae by Members of the Family
 Enterobacteriaceae Isolated from the Human Gastrointestinal Tract
 Infect. Immun. 72 (7), 4151-4158 (2003)
 JOURNAL 2 (bases 1 to 2920)
 AUTHORS Romling, U.
 TITLE Direct Submission
 JOURNAL Submitted (12-NOV-2002) Romling U., Microbiology and Tumorbiology
 Center, Karolinska Institute, Box 280, S-17177 Stockholm, SWEDEN
 FEATURES
 source
 1..2920
 /organism="Citrobacter freundii"
 /mol_type="genomic DNA"
 /isolate="Pec4"
 /isolation_source="faeces"
 /specific_host="Homo sapiens"
 /db_xref="taxon:546"
 /country="Germany"
 /gene="csb"
 complement(211..861)
 /gene="csd"
 complement(211..861)
 /function="necessary for transcription of the csbA
 operon"
 /codon_start=1
 /transl_table=1
 /product="putative regulatory protein"
 /protein_id="CAD56673.1"
 /db_xref="GI:31790496"
 /translation="MREYVSHIGHLTLLITKSLQATALLQKLSLITGKLNQ
 RSLDISSSCIVLDMKADKLIHWQDLNRKNNITLLINTPDYPIINMP
 HINCFYATDEORVVGLOGVGEYSOKLASYLIHSGNRYNSTESALLTHRE
 KEIINKRIGASNIETARSLEISNTVKTLYLFFKIAVKORTQAVSWANDNLR"
 1624..2082
 /gene="csb"
 1624..2082
 /gene="csb"
 /function="nucleator"
 /codon_start=1
 /transl_table=1
 /product="nucleation component of curlin monomers"
 /protein_id="CAD56674.1"
 /db_xref="GI:31790497"

ORIGIN

Query Match 52.0%; Score 237.2; DB 1; Length 2920;
 Best Local Similarity 71.7%; Pred. No. 3.5e-57;
 Matches 327; Conservative 0; Mismatches 123; Indels 6; Gaps 1;

/translation="MKRLIFMFMETMLGVIAAATTSYDLANSEYFAVNELEKSP
 NQAAIQVGTAVNANRTOGSKLSVTSOEGSGNARAKTQTSYNPAVTDQSGSSND
 ASIRQGVNPAVNIIOKSGSNKANITQYGTQKTAVVVQRSGVAIVTQR"
 2123..2572
 /gene="csb"
 2123..2572
 /gene="csb"
 2123..2572
 /function="curli subunit"
 /codon_start=1
 /transl_table=1
 /product="curlin-csb protein"
 /protein_id="CAD56675.1"
 /db_xref="GI:31790498"
 /translation="MKRLVKAAPAAIVSGSALAGVVPQGNHHGGSGNYPDSLS
 IYQVSGNSNANLQSDARKSDVITTHGKNGAVVQGGADDSITSLKQTFONSATID
 QMAKADISVTFGRNALVNTASDSNVLIQVGFNNATANQH"

QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATGTAAGTTTCGCGAGTCTGCT 60
 DB 2123 ATGAACCTTTAAAGTGGCAGCATTCGACGAATGTAAGTTTCGCGAGTCTGCT 2182
 QY 61 GCGGTGCTTCAATGCGGCGCGCGGTATCATACGCGCGCGCAATAGTCCGCG 120
 DB 2183 GGTGTGCTTCCGCAATGCGCGCGCGCGGTATCATACGCGCGCGCAATAGTCCGCG 2236
 QY 121 CCGGATCAACGTTGACATTTATGATGATGATGATGATGATGATGATGATGATGAT 2236
 DB 2237 CCGGATCAACGTTGACATTTATGATGATGATGATGATGATGATGATGATGATGAT 2236
 QY 181 AGCGATGCCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2297
 DB 2297 AGCGATGCCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2356
 QY 241 GCGGCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2416
 DB 2357 GGTGCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2416
 QY 301 GCCACCATGACGACGCTGAGAGCTTAAACTCCGATTTACTGTCCGCAATACGGGGT 360
 DB 2417 GCCACCATGACGACGCTGAGAGCTTAAACTCCGATTTACTGTCCGCAATACGGGGT 2476
 QY 361 AATACGCGCGCGCTGCTTATATCAACCGCATCTGATTCAGCGTAAATGGTCCGTCAGATT 420
 DB 2477 GGCACAGGCTGCGAGTTCAGACCTGACCTTAATCTCCGTCACAGTCACTGAGGTT 2536
 QY 421 GGTTTGGCAACAACGCGCAGCTTACAGATTTAA 456
 DB 2537 GGGTTTGGTACAAACGCGACCGCTCATCAGTACTTA 2572

RESULT 14

A814811 456 bp DNA linear PAT 05-DEC-2003
 LOCUS Sequence 15 from Patent WO03064446.
 DEFINITION A814811
 ACCESSION A814811
 VERSION A814811.1 GI:39104001
 KEYWORDS
 SOURCE
 ORGANISM

Escherichia coli
 Escherichia coli
 Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 Enterobacteriaceae; Escherichia.

REFERENCE 1 Bioerck, L., Olsen, A., Wikstroem, M. and Herwald, H.
 AUTHORS Peptides
 TITLE Patent: WO 03064446-A 15 07-AUG-2003;
 JOURNAL Hansa Medical Research Aktiebolag (SE)

FEATURES
 source
 1..456
 Location/Qualifiers

CDS

```

/organism="Escherichia coli"
/mol_type="unassigned DNA"
/db_xref="taxon:562"
1..456
/feature="unassigned protein product"
/codon_start=1
/translation_table=11
/protein_id="CAE83541.1"
/db_xref="GI:39104002"
/translation="MKLIKVAIAIVSGSALAGVPOYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLITTHGGGNGADVQGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTVAKQFGGNGAAVDQTSNSVNVTVQFGNNATAHOY"

```

ORIGIN

```

Query Match      51.9%; Score 236.8; DB 6; Length 456;
Best Local Similarity 70.0%; Pred. No. 4.1e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTGCTGCTGCA 60
QY 61 GCGCTGTTCCAAATGGGCGCGCGGTAATCAATACGCGCGGCAATAGTTCCGCGC 120
DB 61 GGTGTGTTCCCTCAGTACGCGCGCGGTAATCAATACGCGCGGCAATAGTTCCGCGC 120
QY 121 CCGGACTCAAGCTTATGATTAATGATACGCTTCCGCTAAGCTGCGCTGCTGCA 180
DB 121 CCAAAATTCGAGCTAAACATTTACAGTACGCTGCGGTAATCTGCACTTCTGCA 180
QY 181 AGCGATGCCGCTAAATATGATCAGCTGGTACCGCTGTTGTTACCATGAATGCAAT 240
DB 181 ACTGATGCCGCTAAATATGATCAGCTGGTACCGCTGTTGTTACCATGAATGCAAT 240
QY 241 GCAAGCCAGGCTGCGGTAATGATTAATGATGATGATGATGATGATGATGATGAT 300
DB 241 GTTGTGTTCCCTCAGTACGCGCGGTAATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACATCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 301 GCTACTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AATAAGCCGCGCTGTTAATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 GCGAAGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTGTTGGCAACACGCGCGCTAATGATGATGATGATGATGATGATGATGAT 456
DB 421 GGTGTTGGTAAACACGCGCGCTAATGATGATGATGATGATGATGATGATGAT 456

```

RESULT 15

```

ECOCSSGA 648 bp DNA linear BCT 13-JUL-1993
DEFINITION Escherichia coli curlin subunit (csgA) gene, complete cds.
ACCESSION U04979
VERSION U04979.1 GI:290424
KEYWORDS csgA gene; curli organelle; fibronectin-binding protein.
SOURCE Escherichia coli
ORGANISM Escherichia coli
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Escherichia.
1 (bases 1 to 648)
REFERENCE
AUTHORS Olsen, A., Arngvald, A., Hammar, M., Sukhopolvi, S. and Normark, S.
TITLE The RpoD sigma factor relieves H-NS-mediated transcriptional
repression of csgA, the subunit gene of fibronectin-binding curli
in Escherichia coli
JOURNAL Mol. Microbiol. 7 (4), 523-536 (1993)
MEDLINE 93311294
COMMENT
On Jun 11, 1993 this sequence version replaced gi:145630.
Original source text: Escherichia coli (sub_strain W3110, strain
K-12) (library: Kohara) DNA.

```

FEATURES

```

source
location/Qualifiers
1..648
/organism="Escherichia coli"
/mol_type="genomic DNA"
/strain="K-12"
/sub_strain="W3110"
/db_xref="taxon:562"
/feature="unassigned protein product"
/translation_table=11
/translation="MKLIKVAIAIVSGSALAGVPOYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLITTHGGGNGADVQGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTVAKQFGGNGAAVDQTSNSVNVTVQFGNNATAHOY"
83..538
/gene="csgA"
/feature="precursor protein"
/codon_start=1
/evidence=experimental
/translation_table=11
/product="curlin subunit"
/protein_id="AA23616.1"
/db_xref="GI:290425"
/translation="MKLIKVAIAIVSGSALAGVPOYGGGNGHGGGNSGPNSE
LNIYQYGGNSALALQTDARNSDLITTHGGGNGADVQGSDDSSIDLTRGFGNSAT
LDQWNGKNSMTVAKQFGGNGAAVDQTSNSVNVTVQFGNNATAHOY"
83..142
/gene="csgA"
/evidence=experimental
143..535
/gene="csgA"
/product="curlin subunit"
/evidence=experimental

```

ORIGIN

```

Query Match      51.9%; Score 236.8; DB 1; Length 648;
Best Local Similarity 70.0%; Pred. No. 4.1e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTGCTGCT 60
DB 83 ATGAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTCTGCGAGTCTGCTGCTGCA 142
QY 61 GCGCTGTTCCAAATGGGCGCGCGGTAATCAATACGCGCGGCAATAGTTCCGCGC 120
DB 143 GGTGTGTTCCCTCAGTACGCGCGGTAATCAATACGCGCGGCAATAGTTCCGCGC 120
QY 121 CCGGACTCAAGCTTATGATTAATGATACGCTTCCGCTAAGCTGCGCTGCTGCA 180
DB 121 CCAAAATTCGAGCTAAACATTTACAGTACGCTGCGGTAATCTGCACTTCTGCA 180
QY 181 AGCGATGCCGCTAAATATGATCAGCTGGTACCGCTGTTGTTACCATGAATGCAAT 240
DB 203 CCAAAATTCGAGCTAAACATTTACAGTACGCTGCGGTAATCTGCACTTCTGCA 262
QY 241 GCAAGCCAGGCTGCGGTAATGATTAATGATGATGATGATGATGATGATGATGAT 300
DB 263 ACTGATGCCGCTAAATATGATCAGCTGGTACCGCTGTTGTTACCATGAATGCAAT 322
QY 301 GCCACATCGACGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
DB 383 GCTACTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442
QY 361 AATAAGCCGCGCTGTTAATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 443 GCGAAGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
QY 421 GGTGTTGGCAACACGCGCGCTAATGATGATGATGATGATGATGATGATGAT 456
DB 503 GGTGTTGGTAAACACGCGCGCTAATGATGATGATGATGATGATGATGATGAT 538

```

Search completed: March 15, 2004, 22:50:17
 Job time : 1965.17 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 12:03:05 ; Search time 244.584 Seconds
(without alignments)
7920.305 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456
Sequence: 1 atgaacttcaaaagtcgc.....ccaagcctacacgtrataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

- 1: geneseqn19808:*
- 2: geneseqn19908:*
- 3: geneseqn20008:*
- 4: geneseqn20018:*
- 5: geneseqn20028:*
- 6: geneseqn20038:*
- 7: geneseqn20048:*
- 8: geneseqn20058:*
- 9: geneseqn20068:*
- 10: geneseqn20078:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	3	AAC64626
2	412.8	90.5	456	2	AAQ87467
3	412.8	90.5	456	2	AAQ74142
4	412.8	90.5	456	2	AAQ74142
5	385.6	84.6	456	3	AAC64629
6	363.2	79.6	456	3	AAC64625
7	361.6	79.3	456	3	AAC64628
8	361.6	79.3	456	3	AAC64622
9	361.6	79.3	456	3	AAC64623
10	360	78.9	456	3	AAC64630
11	360	78.9	456	3	AAC64624
12	358.4	78.6	456	3	AAC64627
13	358.4	77.9	456	3	AAC64631
14	288.6	63.3	361	2	AAQ73066
15	288.6	63.3	361	2	AAQ74141
16	236.8	51.9	456	3	AAC64619
17	236.8	51.9	456	3	AAC64619
18	224.4	49.2	646	2	AAQ62647
19	162.4	35.6	369	2	AAQ62646
20	51.2	11.2	100	7	ACD68807
21	50	11.0	78	3	AAC64609
22	50	11.0	78	3	AAC64610
23	48.2	10.6	78	9	ACF36151

24	48	10.5	48	3	AAC64621	AAC64621
25	45	9.9	78	3	AAC64606	AAC64606
26	44.8	9.8	72	3	ACF36148	ACF36148
27	43.4	9.5	78	3	AAC64605	AAC64605
28	43.2	9.5	48	3	AAC64616	AAC64616
29	42.4	9.3	100	7	ACD68808	ACD68808
30	42.2	9.2	2000	7	ADA71938	ADA71938
31	42	9.2	348980	6	ABQ81847	ABQ81847
32	34	7.5	1344	2	AAQ09007	AAQ09007
33	34	7.5	1344	3	AAQ00669	AAQ00669
34	34	7.5	1344	3	AAQ14939	AAQ14939
35	34	7.5	1344	6	AAQ14133	AAQ14133
36	34	7.5	1344	6	AAQ29125	AAQ29125
37	34	7.5	1344	6	AAQ27016	AAQ27016
38	34	7.5	1344	6	ABQ51711	ABQ51711
39	33	7.2	110000	6	ABA90521	ABA90521
40	32.8	7.2	3562	4	ABL06749	ABL06749
41	32.8	7.2	24081	7	AAQ54223	AAQ54223
42	32.8	7.2	53101	7	AAQ54217	AAQ54217
43	32.6	7.1	100	7	ACD68809	ACD68809
44	32.6	7.1	2553	3	AAA09500	AAA09500
45	32.4	7.1	901	4	ABL09481	ABL09481

ALIGNMENTS

RESULT 1
AAC64626
ID AAC64626 standard; DNA; 456 BP.

AC AAC64626;

DT 26-FEB-2001 (first entry)

DE Agfa::PT3#5 DNA sequence SEQ ID NO:19.

XX Salmoneila; agfa; chromosomal gene replacement; fimbriin; epitope;
XX vaccine; immune response; immunogen; ds.

OS Salmoneila enteritidis.

OS Escherichia coli.

OS Synthetic.

PM WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

PA (UYVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay WW;

XX WPI; 2000-672631/65.

DR P-PSDB; AAB36350.

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence

PT which encodes foreign epitope or antigen, expresses recombinant Agfa

PT protein useful for eliciting immune response in animal.

PS Disclosure; Page 137; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA

CC sequence which encodes a foreign epitope or antigen. Also described are:

CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended

CC assembly system of strains of Salmoneila, Escherichia coli and

CC Enterobacteriaceae for the production of fimbriae comprising recombinant

CC Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)

CC directing recombination of a recombinant gene into the chromosome of the

homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fibrillar presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fibrillar protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fibrillar subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fibrillae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 116 A; 111 C; 120 G; 109 T; 0 U; 0 Other;

Query Match 100.0%; Score 456; DB 3; Length 456;

Best Local Similarity 100.0%; Pred. No. 2.2e-139; Mismatches 0; Indels 0; Gaps 0;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGAGTGTCTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGAGTGTCTGCT 60
 QY 61 GGGGTGTTCCAAATGGGGCGGGCGGTAATCATACGGCGCGCAATAGTCCGGC 120
 DB 61 GGGGTGTTCCAAATGGGGCGGGCGGTAATCATACGGCGCGCAATAGTCCGGC 120
 QY 121 CCGGACTCAACGTTAGCATTTATCAGTGGTCCGTAACGCGGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTAGCATTTATCAGTGGTCCGTAACGCGGCTGCTCTGCA 180
 QY 181 AGCGATGCCCGTAATATGATCAGTGGTCCGTAACGCGGCTGCTCTGCA 240
 DB 181 AGCGATGCCCGTAATATGATCAGTGGTCCGTAACGCGGCTGCTCTGCA 240
 QY 241 GCAGGCCAGGGTGGGATTAATGATCTAATGAACTGCAATGCTGTTTCAAAATAT 300
 DB 241 GCAGGCCAGGGTGGGATTAATGATCTAATGAACTGCAATGCTGTTTCAAAATAT 300
 QY 301 GCCACATCGACCAAGTGAACGCTAAAACTCCGATTAATCTGCGCAATACGGCGGT 360
 DB 301 GCCACATCGACCAAGTGAACGCTAAAACTCCGATTAATCTGCGCAATACGGCGGT 360
 QY 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
 DB 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
 QY 421 GGTITGGCAACACGCGCAGGCTAACAGATATTA 456
 DB 421 GGTITGGCAACACGCGCAGGCTAACAGATATTA 456

RESULT 2

AA087467
 ID AA087467 standard; DNA; 456 BP.

XX AA087467;

XX AC

XX 25-MAR-2003 (revised)

DT 26-JUN-1995 (first entry)

XX XX

XX Agfa sequence.

XX *Salmonella*; Agfa; vaccine; genetic immunization; ds.

XX OS

XX KM

XX XX

XX

Key Location/Qualifiers
 CDS 1..454
 /feature="a"
 /note="Agfa"

W09425598-A2.

10-NOV-1994.

26-APR-1994; 94NO-IB000207.

26-APR-1993; 93US-00054452.

(UVI-) UNIV VICTORIA INNOVATION & DEV CORP.

(KING/) KING J.

Key MW, Clouthier SC, Doran JL;

WPI; 1994-358275/44.

P-PSDB; AAR74625.

Eliciting an immune response to *Salmonella* - using attenuated *Salmonella* strains, vector constructs, or compans. conf. fibrillar type proteins.

Disclosure; Fig 7B; 95pp; English.

The DNA encodes the *Salmonella* Agfa protein. The DNA and isolated proteins are used in genetic immunization and vaccine compositions, respectively, to elicit an immune response to *Salmonella* in animals (e.g. food producing animals) and humans. (Updated on 25-MAR-2003 to correct PN field.)

Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 2; Length 456;

Best Local Similarity 94.1%; Pred. No. 3.6e-125;

Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGAGTGTCTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGAGCATTCGAGCAATGCTAGTTCTGGAGTGTCTGCT 60
 QY 61 GGGGTGTTCCAAATGGGGCGGGCGGTAATCATACGGCGCGCAATAGTCCGGC 120
 DB 61 GGGGTGTTCCAAATGGGGCGGGCGGTAATCATACGGCGCGCAATAGTCCGGC 120
 QY 121 CCGGACTCAACGTTAGCATTTATCAGTGGTCCGTAACGCGGCTGCTCTGCA 180
 DB 121 CCGGACTCAACGTTAGCATTTATCAGTGGTCCGTAACGCGGCTGCTCTGCA 180
 QY 181 AGCGATGCCCGTAATATGATCAGTGGTCCGTAACGCGGCTGCTCTGCA 240
 DB 181 AGCGATGCCCGTAATATGATCAGTGGTCCGTAACGCGGCTGCTCTGCA 240
 QY 241 GCAGGCCAGGGTGGGATTAATGATCTAATGAACTGCAATGCTGTTTCAAAATAT 300
 DB 241 GCAGGCCAGGGTGGGATTAATGATCTAATGAACTGCAATGCTGTTTCAAAATAT 300
 QY 301 GCCACATCGACCAAGTGAACGCTAAAACTCCGATTAATCTGCGCAATACGGCGGT 360
 DB 301 GCCACATCGACCAAGTGAACGCTAAAACTCCGATTAATCTGCGCAATACGGCGGT 360
 QY 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
 DB 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
 QY 421 GGTITGGCAACACGCGCAGGCTAACAGATATTA 456
 DB 421 GGTITGGCAACACGCGCAGGCTAACAGATATTA 456

RESULT 3

AA174142

ID	AAT74142standard; DNA; 456 BP.
XX	
AC	AAT74142;
XX	
DT	25-MAR-2003 (revised)
DT	29-SEP-1997 (first entry)
XX	
DE	Salmonella enteritidis 27655-3b agfa gene.
KX	Enteropathogenic bacteria; enterobacteria; S.enteritidis; antibody; ds.
XX	
OS	Salmonella enteritidis.
XX	
PH	Key Location/Qualifiers
FT	CDS 1..456
FT	/tag= a
FT	/label= agfa gene fragment
FT	/trans_except= (pos:367..369,aa:Pro)
XX	
PN	US5635617-A.
PD	
XX	
XX	03-JUN-1997.
XX	
Pf	26-APR-1994; 94US-00233788.
XX	
PR	26-APR-1993; 93US-00054452.
XX	
PA	(UUVI-) UNIV VICTORIA INNOVATION & DEV CORP.
XX	
PI	Collinson SK, Kay WW, Doran JL;
DR	WPI; 1997-309886/28.
DR	P-PSDB; AAM23570.
XX	
PT	Isolated Salmonella gene agfa - used for diagnosis of Salmonella or
PT	enteropathogenic bacteria of the Enterobacteria family.
XX	
XX	Claim 1; Col 19-112; 85pp; English.
XX	
CC	The present sequence represents an isolated agfa gene derived from
CC	Salmonella enteritidis 27655-3b. The nucleic acid can be used to provide
CC	diagnostic assays for Salmonella and/or enteropathogenic bacteria of the
CC	family Enterobacteria. It can also be used to provide proteins and
CC	antibodies which can be used for assays. The nucleic acid sequence can be
CC	used to provide probes or primers which can specifically hybridise to
CC	nucleic acid molecules from greater than 99% of Salmonella strains that
CC	are pathogenic to warm-blooded animals relative to nucleic acid molecules
CC	from virtually all other microbial organisms. (Updated on 25-MAR-2003 to
CC	correct Pf field.)
SQ	Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;
Query Match	90.5%; Score 412.8; DB 2; Length 456;
Best Local Similarity	94.1%; Pred. No. 3.6e-125;
Matches 429; Conservative	0; Mismatches 27; Indels 0; Gaps 0;
OY	1 ATGAAACTTTTAAAAGTGCGACATTGGCAGCAATCGTAGTTCTGGCAGTCCTGACT 60
Db	1 ATGAACCTTTTAAAAGTGCGACATTGGCAGCAATCGTAGTTCTGGCAGTCCTGACT 60
OY	61 GGCGTCGTTCACAATGGGGGGCGCGCGCATATCATTAACGGCGCGGCAGATAGTCCGCC 120
Db	61 GGCGTCGTTCACAATGGGGGGCGCGCGCATATCATTAACGGCGCGGCAGATAGTCCGCC 120
OY	121 CCGGACTCAAGTTGAGCATTTATAGTAGCGGTTCCGCTAACGCTGCGCTTGCTGCAA 180
Db	121 CCGGACTCAAGTTGAGCATTTATAGTAGCGGTTCCGCTAACGCTGCGCTTGCTGCAA 180
OY	181 AGCGATGCCCGTAATAATATGATCAGCTGGTTACCCTGTTGTTATCCCATGAATGGCACAT 240
Db	181 AGCGATGCCCGTAATAATATGTAACACACCATTAACCGACGCGTTAATGTAACGGCGCCAT 240
OY	241 GCAGCGCAGGGTGGCGATTAATAGTACTTGAATGAACTGACTCAGAAATGTTTCAGAATAAT 300

Dd	241	GTAGGCGCAGGGGCGGATTAATAGTACTATTTGAACGACTCAAGAAATGGTTTCAGAAATTAAT	300
Oy	301	GGCACCATTCAGACCACTGGAGAACGCTAAAACTCCGATATTACTGTCCGCCAATTCGGCCGGT	360
Dd	301	GGCACCATTCAGACCACTGGAGAACGCTAAAACTCCGATATTACTGTCCGCCAATTCGGCCGGT	360
Oy	361	AATAACGCCCGCGCTGTGTAATCAGACCCGCATCTGATTCAGCGCTAAATGGTGGCTCAGGTT	420
Dd	361	AATAACGCCCGCGCTGTGTAATCAGACCCGCATCTGATTCAGCGCTAAATGGTGGCTCAGGTT	420
Oy	421	GGTTTGGCACAACAGCCACGGCTTAACCAAGTATTAA	456
Dd	421	GGTTTGGCACAACAGCCACGGCTTAACCAAGTATTAA	456
RESUT 4			
ID	AAC64617	standard; DNA; 456 BP.	
AC	AAC64617;		
XX			
XX	26-FEB-2001	(first entry)	
DT			
DE	Salmonella enteritidis Agfa DNA sequence SEQ ID NO.1.		
XX			
XX	Salmonella agfa; chromosomal gene replacement; fimbriin; epitope;		
KW	vaccine; immune response; immunogen; ds.		
XX			
OS	Salmonella enteritidis.		
PN	WO200060102-A2.		
PD	12-OCT-2000.		
XX			
XX	05-APR-2000; 2000WO-CA000356.		
PF			
PR	05-APR-1999; 99US-0127888P.		
XX			
XX	(UUYV-) UNIV VICTORIA.		
PA			
P1	White AP, Doran JL, Collison SK, Kay WJ,		
XX	WPI; 2000-672631/65.		
DR	P-PSDB; AAB36341.		
XX			
PT	Recombinant agfa gene having a segment replaced by a foreign DNA sequence		
PT	which encodes foreign epitope or antigen, expresses recombinant Agfa		
XX	protein useful for eliciting immune response in animal.		
PS	Disclosure; Page 134; 139pp; English.		
XX			
XX	The present invention describes a recombinant agfa gene (1) where a		
CC	segment of the gene has been replaced by a segment of a foreign DNA		
CC	sequence which encodes a foreign epitope or antigen. Also described are:		
CC	(1) use of thin aggregative fimbriae (SEF17/TAI) nucleation depended		
CC	assembly system of strains of Salmonella, Escherichia coli and		
CC	Enterobacteriaceae for the production of fimbriae comprising recombinant		
CC	Agfa, CsgA and Agfa-homologue fimbriin subunits, respectively; (2)		
CC	directing recombination of a recombinant gene into the chromosome of the		
CC	homologous species; (3) directing recombination of a recombinant gene		
CC	back into the chromosome of the homologous species, replacing the native		
CC	copy of that gene; and (4) eliciting an immune response in an animal,		
CC	comprising separating an amino acid polymer comprising a recombinant Agfa		
CC	protein containing a replacement segment or segments of foreign amino		
CC	acid sequence or sequences grown on a Salmonella, E. coli or		
CC	Enterobacteriaceae host cell, from the host cell and introducing the		
CC	polymer into the animal in conjunction with a carrier or diluent. (1) is		
CC	useful for the expression of recombinant Agfa protein which is useful for		
CC	eliciting an immune response in an animal. In a fimbriin presentation		
CC	system the heterologous antigens are presented in high numbers (up to		
CC	500,000 copies/cell), the hybrid fimbriin protein possesses both the		
CC	immunogenicity and adhesion properties relevant for an efficient live		

CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 456 BP; 117 A; 112 C; 122 G; 105 T; 0 U; 0 Other;

Query Match 90.5%; Score 412.8; DB 3; Length 456;
 Best Local Similarity 94.1%; Pred. No. 3.6e-125;
 Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGAGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGAGAGCTCTGCT 60
QY 61 GCGGTGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
DB 61 GCGGTGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGGTAAACGCTGCTTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGGTAAACGCTGCTTGCAA 180
QY 181 AGCGATGCCGCTTAAATATGATCAGCTGTTACCGGTGTTTACCATGAATGCGACAT 240
DB 181 AGCGATGCCGCTTAAATATGAAACGACCATTAACGAGCGGTATGTTAACGCGCCGAT 240
QY 241 GAGGCGCAGGTTGGGATTAATAGTATGAACTGATGACGATCGTAAAGTTCGAAATAT 300
DB 241 GTAGGCGCAGGTTGGGATTAATAGTATGAACTGATGACGATCGTAAAGTTCGAAATAT 300
QY 301 GCCACATTCGACGCTGTAAGCTTAAACCTCGATATTAAGTGTGGCCATACGGCGGT 360
DB 301 GCCACATTCGACGCTGTAAGCTTAAACCTCGATATTAAGTGTGGCCATACGGCGGT 360
QY 361 AATAACCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTTATGTTGCTGAGTT 420
DB 361 AATAACCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTTATGTTGCTGAGTT 420
QY 421 GGTTTGGCAACAAACCGCGCTAACGATATTA 456
DB 421 GGTTTGGCAACAAACCGCGCTAACGATATTA 456

```

RESULT 5
 AAC64629
 ID AAC64629 standard; DNA; 456 BP.
 XX
 AC AAC64629;
 XX
 DT 26-FEB-2001 (first entry)
 XX
 DE Agfa: PT3#8 DNA sequence SEQ ID NO:25.
 XX
 KW Salmonella; agfa: chromosomal gene replacement; fimbria; epitope;
 XX vaccine; immune response; immunogen; ds.
 XX
 OS Salmonella enteritidis.
 OS Escherichia coli.
 OS Synthetic.
 OS
 PN WO200060102-A2.
 XX
 PD 12-OCT-2000.
 XX
 PF 05-APR-2000; 2000WO-CA000356.
 XX
 PR 05-APR-1999; 99US-0127888P.
 XX
 PA (UVI-) UNIV VICTORIA.
 XX
 PI White AP, Doran JL, Collison SK, Kay WM;

XX
 DR WPI; 2000-672631/65.
 DR P-PSDB; AAB36353.
 XX

PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 PS Disclosure; Page 138; 139pp; English.

The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/7AP) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

SO Sequence 456 BP; 114 A; 108 C; 123 G; 111 T; 0 U; 0 Other;

Query Match 84.6%; Score 385.6; DB 3; Length 456;
 Best Local Similarity 90.4%; Pred. No. 3.1e-116;
 Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

```

QY 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGAGAGCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGACGATTCGACGATCGTAGTTCTGGAGAGCTCTGCT 60
QY 61 GCGGTGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
DB 61 GCGGTGTTCCACATGGGGCGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGGTAAACGCTGCTTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGCTTCGGTAAACGCTGCTTGCAA 180
QY 181 AGCGATGCCGCTTAAATATGATCAGCTGTTACCGGTGTTTACCATGAATGCGACAT 240
DB 181 AGCGATGCCGCTTAAATATGATCAGCTGTTACCGGTGTTTACCATGAATGCGACAT 240
QY 241 GAGGCGCAGGTTGGGATTAATAGTATGAACTGATGACGATCGTAAAGTTCGAAATAT 300
DB 241 GTAGGCGCAGGTTGGGATTAATAGTATGAACTGATGACGATCGTAAAGTTCGAAATAT 300
QY 301 GCCACATTCGACGCTGTAAGCTTAAACCTCGATATTAAGTGTGGCCATACGGCGGT 360
DB 301 GCCACATTCGACGCTGTAAGCTTAAACCTCGATATTAAGTGTGGCCATACGGCGGT 360
QY 361 AATAACCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTTATGTTGCTGAGTT 420
DB 361 AATAACCGCGGCTGTTATATCAGACCGCATCGATTCCAGGTTATGTTGCTGAGTT 420
QY 421 GGTTTGGCAACAAACCGCGCTAACGATATTA 456
DB 421 GGTTTGGCAACAAACCGCGCTAACGATATTA 456

```

Db 421 GGTTTGGCAACGCGCTAACGATATTA 456

RESULT 6
AAC64625
ID AAC64625 standard; DNA; 456 BP.
XX
AC AAC64625;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#4 DNA sequence SEQ ID NO:17.
XX
KM Salmone11a; agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmone11a enteritidis.
OS Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
DR P-PSDB; AAB36349.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 136; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF17/Tra) nucleation depended assembly system of strains of Salmone11a, Escherichia coli and Enterobacteriaceae for the production of fimbrial subunits, respectively; (2) Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (3) the directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a Salmone11a, E. coli or Enterobacteriaceae host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (I) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbrial protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 120 A; 110 C; 122 G; 104 T; 0 U; 0 Other;

Query Match 79.6%; Score 363.2; DB 3; Length 456;
Best Local Similarity 87.3%; Pred. No. 7.3e-109;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGACAGATTCCAGCATCTGATGTTCTGCGAGTCTTGCT 60
Db 1 ATGAACCTTTTAAAGTGGACAGATTCCAGCATCTGATGTTCTGCGAGTCTTGCT 60
QY 61 GCGCGTGTTCACATATGAGGCGGCGGCGGTATCATACGCGGCGCAATAGTCCGCG 120
Db 61 GCGCGTGTTCACATATGAGGCGGCGGCGGTATCATACGCGGCGCAATAGTCCGCG 120
QY 121 CCGGACTGCAACGTGAGCATTTATCAGTACGCGTTCGCTAACGCTGCGCTTGCGAA 180
Db 121 CCGGACTGATGATCAGCTGTATACCGGTGTGTATCCCATGAAATGCGCATGCGAA 180
QY 181 AGCGATGCCCGTAAATATGATCAGTGTATCCCGTGTGTTTACCATGAAATGCGCAT 240
Db 181 AGCGATGCCCGTAAATATGAAACGCACTTCCAGACGCGGTATGATGAAACGCGCAT 240
QY 241 GCAGGCGAGGCTGGGATATATGATATGATATGATGATGATGATGATGATGATGAT 300
Db 241 GTAGGCGAGGCTGGGATATATGATATGATATGATGATGATGATGATGATGAT 300
QY 301 GCCACCATCGACAGTGGAAAGCTTAAATATGATGATGATGATGATGATGATGAT 360
Db 301 GCCACCATCGACAGTGGAAAGCTTAAATATGATGATGATGATGATGATGATGAT 360
QY 361 AATAACGCGCGCGTGTATATCAGACCGCATCTGATTCACGCTAATGCTGCAAGTT 420
Db 361 AATAACGCGCGCGTGTATATCAGACCGCATCTGATTCACGCTAATGCTGCAAGTT 420
QY 421 GGTTTGGCAACAGCGCGCTAACGATATTA 456
Db 421 GGTTTGGCAACAGCGCGCTAACGATATTA 456

RESULT 7
AAC64628
ID AAC64628 standard; DNA; 456 BP.
XX
AC AAC64628;
XX
DT 26-FEB-2001 (first entry)
XX
DE Agfa::PT3#7 DNA sequence SEQ ID NO:23.
XX
KM Salmone11a; agfa; chromosomal gene replacement; fimbriin; epitope;
KW vaccine; immune response; immunogen; ds.
XX
OS Salmone11a enteritidis.
OS Escherichia coli.
XX Synthetic.
XX WO200060102-A2.
XX
PD 12-OCT-2000.
XX
PF 05-APR-2000; 2000MO-CA000356.
XX
PR 05-APR-1999; 99US-0127888P.
XX
PA (UYVI-) UNIV VICTORIA.
XX
PI White AP, Doran JL, Collison SK, Kay WW;
XX WPI: 2000-672631/65.
DR P-PSDB; AAB36352.
XX
PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
PS Disclosure; Page 137; 139pp; English.

The present invention describes a recombinant agfa gene (I) where a

CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 119 A; 110 C; 120 G; 107 T; 0 U; 0 Other;

XX Query Match 79.3%; Score 361.6; DB 3; Length 456;

XX Best Local Similarity 87.1%; Pred. No. 2.5e-108;

XX Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCGAGAGTCTTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCGAGAGTCTTGCT 60
 QY 61 GCGCTGCTTCCAAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 DB 61 GCGCTGCTTCCAAATGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 QY 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 QY 181 AGCGATGCCCGTAAATGATGATGATGATGATGATGATGATGATGATGAT 240
 DB 181 AGCGATGCCCGTAAATGATGATGATGATGATGATGATGATGATGATGAT 240
 QY 241 GCAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
 DB 241 GCAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
 QY 301 GCCACCATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 DB 301 GCCACCATGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
 QY 361 AATTAACCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 420
 DB 361 AATTAACCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 420
 QY 421 GGTGGGCAACGAGCGGCGGCTAACGATATTTAA 456
 DB 421 GGTGGGCAACGAGCGGCGGCTAACGATATTTAA 456

RESULT 8
 AAC64622
 ID AAC64622 standard; DNA; 456 BP.

XX AAC64622;
 DT 26-FEB-2001 (first entry)

DE Agfa::PT3#1 DNA sequence SEQ ID NO:11.
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KM vaccine; immune response; immunogen; ds.
 XX *Salmonella enteritidis*.
 OS *Escherichia coli*.
 OS Synthetic.
 PM WO200060102-A2.
 XX 12-OCT-2000.
 PD 05-APR-2000; 2000MO-CA000356.
 PF 05-APR-1999; 99US-0127888P.
 PR (UVTI-) UNIV VICTORIA.
 PA White AP, Doran JL, Collison SK, Kay MW;
 PI WPI: 2000-672631/65.
 DR P-PSDB: AAB36346.
 XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX Disclosure; Page 135; 139pp; English.

XX The present invention describes a recombinant agfa gene (I) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/RAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, Caga and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating an amino acid polymer comprising a recombinant Agfa
 CC protein containing a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention

XX Sequence 456 BP; 121 A; 112 C; 118 G; 105 T; 0 U; 0 Other;

XX Query Match 79.3%; Score 361.6; DB 3; Length 456;

XX Best Local Similarity 87.1%; Pred. No. 2.5e-108;

XX Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCGAGAGTCTTGCT 60
 DB 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGAGTTCGAGAGTCTTGCT 60
 QY 61 GCGCTGCTTCCAAATGGGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 DB 61 GCGCTGCTTCCAAATGGGCGCGGTATCATTAACGGCGGCAATAGTTCCGGC 120
 QY 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180
 DB 121 CCGGACTCAAGCTGATGATTAATGATGATGATGATGATGATGATGATGATGAT 180


```
Db 121 CCGACTCAACGTTGAGCATTTATCATGACGGTTCCTGCAACGCTGCGCTTCTCTGCA 180
Oy 181 AGCGATGCGCGTAAATATGATAGCTGGTACCCGTGTTTACCAGTAAGAAATGGCAAT 240
Db 181 AGCGATGCGCGTAAATCTGAAAAGCAATTCACAGACGGGTTATGTAACGGCGCGAT 240
Oy 241 GCAGCCGAGGTCGCGATTAATAGTACTATTGAACTGACTGAAATGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTGAACTGACTGAAATGTTTCAGAAATAT 300
Oy 301 GCCACATGACCACTGGAACGCTTAAAACTCCGATTTACTGTCGCCAATACGGCGGT 360
Db 301 GCCACATGACCACTGGAACGCTTAAAACTCCGATTTACTGTCGCCAATACGGCGGT 360
Oy 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCACGCTAATGTCGTCAGGTT 420
Db 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCACGCTAATGTCGTCAGGTT 420
Oy 421 GGTTCGCAACAAACGCGACGCTAACCGATTTAA 456
Db 421 GCACATGCAACAAACGCGCTAACCGATTTAA 456

RESULT 9
ID AAC64623 standard; DNA; 456 BP.
AC AAC64623;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#2 DNA sequence SEQ ID NO:13.
DE
XX
XX Salmone11a; agfa, chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX Salmone11a enteritidis.
OS
XX Escherichia coli.
OS
XX Synthetic.
XX
XX WO200060102-A2.
PN
XX
XX 12-OCT-2000.
PD
XX
XX 05-APR-2000; 2000WO-CA000356.
PP
XX
XX 05-APR-1999; 99US-0127888P.
PR
XX
XX (UUVI-) UNIV VICTORIA.
PA
XX
XX White AP, Doran JL, Collison SK, Kay WM;
PI
XX WPI; 2000-672631/65.
DR
XX P-PSDB; AAB36347.
XX
XX Recombinant agfa gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant Agfa
PT protein useful for eliciting immune response in animal.
XX
XX Disclosure; Page 136; 139p; English.
XX
XX The present invention describes a recombinant agfa gene (I) where a
XX segment of the gene has been replaced by a segment of a foreign DNA
XX sequence which encodes a foreign epitope or antigen. Also described are:
XX (1) use of thin aggregative fimbriae (SEF1/TFP) nucleation depended
XX assembly system of strains of Salmone11a, Escherichia coli and
XX Enterobacteriaceae for the production of fimbriae comprising recombinant
XX Agfa, CagA and Agfa-homologue fimbriin subunits, respectively; (2)
XX directing recombination of a recombinant gene into the chromosome of the
XX homologous species; (3) directing recombination of a recombinant gene
XX back into the chromosome of the homologous species, replacing the native
XX copy of that gene; and (4) eliciting an immune response in an animal,
XX comprising separating an amino acid polymer comprising a recombinant Agfa
```

```
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a Salmone11a, E. coli or
CC Enterobacteriaceae host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant Agfa protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbriin protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

SQ Sequence 456 BP; 118 A; 109 C; 121 G; 108 T; 0 U; 0 Other;

Query Match 79.3%; Score 361.6; DB 3; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.5e-108;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Oy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGCGAGTCTTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGCGAGTCTTGCT 60
Oy 61 GCGCGTTCGTCACAAATGGCGCGCGGCGGTAATCATATAACGGCGGCAATGTTCCGGC 120
Db 61 GCGCGTTCGTCACAAATGGCGCGCGGCGGTAATCATATAACGGCGGCAATGTTCCGGC 120
Oy 121 CCGGACTCAACGTTGAGCATTTATCATGACGGTTCGCTTAAACGTCGCGCTTCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCATGACGGTTCGCTTAAACGTCGCGCTTCTGCA 180
Oy 181 AGCGATGCGCGTAAATATGATAGCTGGTACCCTGTTGTTACCATGAATGCGACAT 240
Db 181 AGCGATGCGCGTAAATATGATAGCTGGTACCCTGTTGTTACCATGAATGCGACAT 240
Oy 241 GCAGCCGAGGTCGCGATTAATAGTACTATTGAACTGACTGAAATGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTGAACTGACTGAAATGTTTCAGAAATAT 300
Oy 301 GCCACATGACCACTGGAACGCTTAAAACTCCGATTTACTGTCGCCAATACGGCGGT 360
Db 301 GCCACATGACCACTGGAACGCTTAAAACTCCGATTTACTGTCGCCAATACGGCGGT 360
Oy 361 AATAACGCGCGGTGTTAATCAGACCGCATCTGATTCACGCTAATGTCGTCAGGTT 420
Db 361 CTGGTTACCGGTGTTGTTAACCATGAATGACATCAAGCGTAATGTCGTCAGGTT 420
Oy 421 GGTTCGCAACAAACGCGACGCTAACCGATTTAA 456
Db 421 GGTTCGCAACAAACGCGACGCTAACCGATTTAA 456

RESULT 10
ID AAC64630 standard; DNA; 456 BP.
AC AAC64630;
XX
XX 26-FEB-2001 (first entry)
DT
XX
XX Agfa::PT3#9 DNA sequence SEQ ID NO:27.
DE
XX
XX Salmone11a; agfa, chromosomal gene replacement; fimbriin; epitope;
KM vaccine; immune response; immunogen; ds.
XX
XX Salmone11a enteritidis.
OS
XX Escherichia coli.
OS
XX Synthetic.
XX
XX WO200060102-A2.
XX
```

PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 XX
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 XX
 XX White AP, Doran JL, Collison SK, Kay MW;
 XX WPI; 2000-672631/65.
 XX P-PSDB; AAB36354.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 138; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response
 CC against the inserted epitope, and hybrid fimbriae are easy and
 CC inexpensive to purify in large amount. The present sequence is given in
 CC the exemplification of the present invention
 XX
 XX Sequence 456 BP; 115 A; 116 C; 118 G; 107 T; 0 U; 0 Other;
 SQ
 Query Match 78.9%; Score 360; DB 3; Length 456;
 Best Local Similarity 86.8%; Pred. No. 8.3e-108;
 Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
 QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCTGCT 60
 DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATCGTAGTTCTGGCAGTGGCTGCT 60
 QY 61 GCGCGCTTCCACATGAGGCGGCGGGGATATATACGCGCGGCAATAGTTCGCGC 120
 DB 61 GCGCGCTTCCACATGAGGCGGCGGGGATATATACGCGCGGCAATAGTTCGCGC 120
 QY 121 CCGAGCTCAACGTTGAGCATTTACAGTACGTTCCGCTACGCTGCGCTGCTGCA 180
 DB 121 CCGAGCTCAACGTTGAGCATTTACAGTACGTTCCGCTACGCTGCGCTGCTGCA 180
 QY 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGCTGTTTATCCCATGAAATGGCAGAT 240
 DB 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGCTGTTTATCCCATGAAATGGCAGAT 240
 QY 241 GCAGGCGGAGGTGGGATATAGACATTAAGTGAAGTCAATGATGTTTCAAAATAT 300
 DB 241 TATGATCAGGTGTTACCGCTGTTTATCCCATGAAATGGCAGTATTCAGAAATAT 300
 QY 301 GCCACATGACGACGATGGAACGCTAAATCTCGATATTAATGCTGGCCATATACGCGGT 360

DB 301 GCCACATGACGACGATGGAACGCTAAATCTCGATATTAATGCTGGCCATATACGCGGT 360
 QY 361 AATAACGCGCGGCTGTTAAATCAGACCGCATTCGATTCAGCGTAAATGGGCTCAGGTT 420
 DB 361 AATAACGCGCGGCTGTTAAATCAGACCGCATTCGATTCAGCGTAAATGGGCTCAGGTT 420
 QY 421 GCTTTGGCAACACGCGCAGCGCTAACAGTATTA 456
 DB 421 GCTTTGGCAACACGCGCAGCGCTAACAGTATTA 456
 RESULT 11
 AAC64624
 ID AAC64624 standard; DNA; 456 BP.
 XX
 XX AAC64624;
 XX
 XX 26-FEB-2001 (first entry)
 XX
 XX Agfa::PT3#3 DNA sequence SEQ ID NO:15.
 DE
 XX *Salmonella*; agfa; chromosomal gene replacement; fimbrial; epitope;
 KW vaccine; immune response; immunogen; ds.
 XX
 XX *Salmonella enteritidis*.
 OS
 XX *Escherichia coli*.
 OS Synthetic.
 PN WO200060102-A2.
 PD 12-OCT-2000.
 XX
 XX 05-APR-2000; 2000WO-CA000356.
 PF
 XX 05-APR-1999; 99US-0127888P.
 XX
 XX (UYVI-) UNIV VICTORIA.
 PA
 XX White AP, Doran JL, Collison SK, Kay MW;
 PI WPI; 2000-672631/65.
 XX P-PSDB; AAB36348.
 XX
 PT Recombinant agfa gene having a segment replaced by a foreign DNA sequence
 PT which encodes foreign epitope or antigen, expresses recombinant Agfa
 PT protein useful for eliciting immune response in animal.
 XX
 PS Disclosure; Page 136; 139pp; English.
 XX
 CC The present invention describes a recombinant agfa gene (1) where a
 CC segment of the gene has been replaced by a segment of a foreign DNA
 CC sequence which encodes a foreign epitope or antigen. Also described are:
 CC (1) use of thin aggregative fimbriae (SEF17/TAF) nucleation depended
 CC assembly system of strains of *Salmonella*, *Escherichia coli* and
 CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
 CC Agfa, CsgA and Agfa-homologue fimbrial subunits, respectively; (2)
 CC directing recombination of a recombinant gene into the chromosome of the
 CC homologous species; (3) directing recombination of a recombinant gene
 CC back into the chromosome of the homologous species, replacing the native
 CC copy of that gene; and (4) eliciting an immune response in an animal,
 CC comprising separating a replacement segment or segments of foreign amino
 CC acid sequence or sequences grown on a *Salmonella*, *E. coli* or
 CC *Enterobacteriaceae* host cell, from the host cell and introducing the
 CC polymer into the animal in conjunction with a carrier or diluent. (1) is
 CC useful for the expression of recombinant Agfa protein which is useful for
 CC eliciting an immune response in an animal. In a fimbrial presentation
 CC system the heterologous antigens are presented in high numbers (up to
 CC 500,000 copies/cell), the hybrid fimbrial protein possesses both the
 CC immunogenicity and adhesion properties relevant for an efficient live
 CC vaccine, the carrier fimbrial subunit proteins are usually strong
 CC immunogens, which may be important for directing an immune response

CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 118 A; 112 C; 115 G; 111 T; 0 U; 0 Other;

Query Match 78.9%; Score 360; DB 3; Length 456;
Best Local Similarity 86.8%; Pred. No. 8.3e-108;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```

OY 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAAGATGTAAGTTCTGGAGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAAGATGTAAGTTCTGGAGAGTCTTGCT 60
OY 61 GCGCTCTTATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCATTC 120
DB 61 GCGCTCTTATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCATTC 120
OY 121 CCGGACTCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 CCGGACTCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
OY 181 AGCGATGCCGCTAAATATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 240
DB 181 AGCGATGCCGCTAAATATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 240
OY 241 GCGAGCCGAGGCTGCGGATTAATGATCTATGATGATGATGATGATGATGATGATGAT 300
DB 241 GCGAGCCGAGGCTGCGGATTAATGATCTATGATGATGATGATGATGATGATGATGAT 300
OY 301 GCCACATCGACGATGGAAGCGCTTAAACTCCGATTTACTGTCGGGCAATACGGCGGT 360
DB 301 GCCACATCGACGATGGAAGCGCTTAAACTCCGATTTACTGTCGGGCAATACGGCGGT 360
OY 361 AATAAGCCGCGCTGTTAATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 420
DB 361 AATAAGCCGCGCTGTTAATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 420
OY 421 GGTTTGGCAACAAAGCGGCTTAAATGATGATGATGATGATGATGATGATGATGAT 456
DB 421 GGTTTGGCAACAAAGCGGCTTAAATGATGATGATGATGATGATGATGATGATGAT 456

```

RESULT 12

AAC64627
ID AAC64627 standard; DNA; 456 BP.

XX AAC64627;

XX 26-FEB-2001 (first entry)

DE AgfA::PT3#6 DNA sequence SEQ ID NO:21.

XX Salmomella; agfA; chromosomal gene replacement; fimbria; epitope;
KV vaccine; immune response; immunogen; ds.

XX Salmomella enteritidis.

OS Escherichia coli.

OS Synthetic.

XX WO200060102-A2.

PD 12-OCT-2000.

PF 05-APR-2000; 2000WO-CA000356.

PR 05-APR-1999; 99US-0127888P.

XX (UVI-) UNIV VICTORIA.

PI White AP, Doran JL, Collison SK, Kay MW;
XX WPI; 2000-672631/65.

DR P-PSDB; AAB36351.

XX Recombinant agfA gene having a segment replaced by a foreign DNA sequence
PT which encodes foreign epitope or antigen, expresses recombinant AgfA
XX protein useful for eliciting immune response in animal.

XX Disclosure; Page 137; 139pp; English.

CC The present invention describes a recombinant agfA gene (I) where a
CC segment of the gene has been replaced by a segment of a foreign DNA
CC sequence which encodes a foreign epitope or antigen. Also described are:
CC (1) use of thin aggregative fimbriae (SAF17/TF) nucleotide dependent
CC assembly system of strains of *Salmomella*, *Escherichia coli* and
CC *Enterobacteriaceae* for the production of fimbriae comprising recombinant
CC AgfA, CsgA and AgfA-homologue fimbria subunit, respectively; (2)
CC directing recombination of a recombinant gene into the chromosome of the
CC homologous species; (3) directing recombination of a recombinant gene
CC back into the chromosome of the homologous species, replacing the native
CC copy of that gene; and (4) eliciting an immune response in an animal,
CC comprising separating an amino acid polymer comprising a recombinant AgfA
CC protein containing a replacement segment or segments of foreign amino
CC acid sequence or sequences grown on a *Salmomella*, *E. coli* or
CC *Enterobacteriaceae* host cell, from the host cell and introducing the
CC polymer into the animal in conjunction with a carrier or diluent. (I) is
CC useful for the expression of recombinant AgfA protein which is useful for
CC eliciting an immune response in an animal. In a fimbrial presentation
CC system the heterologous antigens are presented in high numbers (up to
CC 500,000 copies/cell), the hybrid fimbria protein possesses both the
CC immunogenicity and adhesion properties relevant for an efficient live
CC vaccine, the carrier fimbrial subunit proteins are usually strong
CC immunogens, which may be important for directing an immune response
CC against the inserted epitope, and hybrid fimbriae are easy and
CC inexpensive to purify in large amount. The present sequence is given in
CC the exemplification of the present invention

XX Sequence 456 BP; 112 A; 113 C; 125 G; 106 T; 0 U; 0 Other;

Query Match 78.6%; Score 358.4; DB 3; Length 456;
Best Local Similarity 86.6%; Pred. No. 2.8e-107;
Matches 395; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

```

OY 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAAGATGTAAGTTCTGGAGAGTCTTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTTGCGAAGATGTAAGTTCTGGAGAGTCTTGCT 60
OY 61 GCGCTCTTATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCATTC 120
DB 61 GCGCTCTTATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCATTC 120
OY 121 CCGGACTCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
DB 121 CCGGACTCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
OY 181 AGCGATGCCGCTAAATATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 240
DB 181 AGCGATGCCGCTAAATATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 240
OY 241 GCGAGCCGAGGCTGCGGATTAATGATCTATGATGATGATGATGATGATGATGATGAT 300
DB 241 GCGAGCCGAGGCTGCGGATTAATGATCTATGATGATGATGATGATGATGATGATGAT 300
OY 301 GCCACATCGACGATGGAAGCGCTTAAACTCCGATTTACTGTCGGGCAATACGGCGGT 360
DB 301 GCCACATCGACGATGGAAGCGCTTAAACTCCGATTTACTGTCGGGCAATACGGCGGT 360
OY 361 AATAAGCCGCGCTGTTAATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 420
DB 361 AATAAGCCGCGCTGTTAATGATCAGCTGTTACCCGCTGTTTACCCATGAAATGGCAGCAT 420
OY 421 GGTTTGGCAACAAAGCGGCTTAAATGATGATGATGATGATGATGATGATGATGAT 456
DB 421 GGTTTGGCAACAAAGCGGCTTAAATGATGATGATGATGATGATGATGATGATGAT 456

```

RESULT 13

AAC64631 standard; DNA; 456 BP.

AAC64631;

26-FEB-2001 (first entry)

Agfa::PT3#10 DNA sequence SEQ ID NO:29.

Salmonella; agfa; chromosomal gene replacement; fimbria; epitope; vaccine; immune response; immunogen; ds.

Salmonella enteritidis.

Escherichia coli.

Synthetic.

MO200060102-A2.

12-OCT-2000.

05-APR-2000; 2000MO-CA000356.

05-APR-1999; 99US-0127888P.

(UYVI-) UNIV VICTORIA.

White AP, Doran JL, Collison SK, Kay WJ,

WPI; 2000-672631/65.

P-PSDB; AAB36355.

Recombinant agfa gene having a segment replaced by a foreign DNA sequence which encodes foreign epitope or antigen, expresses recombinant Agfa protein useful for eliciting immune response in animal.

Disclosure; Page 138; 139P; English.

The present invention describes a recombinant agfa gene (1) where a segment of the gene has been replaced by a segment of a foreign DNA sequence which encodes a foreign epitope or antigen. Also described are: (1) use of thin aggregative fimbriae (SEF1/TAFF) nucleation depended assembly system of strains of *Salmonella*, *Escherichia coli* and *Enterobacteriaceae* for the production of fimbriae comprising recombinant Agfa, CsgA and Agfa-homologue fimbria subunits, respectively; (2) directing recombination of a recombinant gene into the chromosome of the homologous species; (3) directing recombination of a recombinant gene back into the chromosome of the homologous species, replacing the native copy of that gene; and (4) eliciting an immune response in an animal, comprising separating an amino acid polymer comprising a recombinant Agfa protein containing a replacement segment or segments of foreign amino acid sequence or sequences grown on a *Salmonella*, *E. coli* or *Enterobacteriaceae* host cell, from the host cell and introducing the polymer into the animal in conjunction with a carrier or diluent. (1) is useful for the expression of recombinant Agfa protein which is useful for eliciting an immune response in an animal. In a fimbrial presentation system the heterologous antigens are presented in high numbers (up to 500,000 copies/cell), the hybrid fimbria protein possesses both the immunogenicity and adhesion properties relevant for an efficient live vaccine, the carrier fimbrial subunit proteins are usually strong immunogens, which may be important for directing an immune response against the inserted epitope, and hybrid fimbriae are easy and inexpensive to purify in large amount. The present sequence is given in the exemplification of the present invention

Sequence 456 BP; 113 A; 110 C; 124 G; 109 T; 0 U; 0 Other;

Query Match 77.9%; Score 355.2; DB 3; Length 456;

Best Local Similarity 86.2%; Pred. No. 3.1e-106;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

1 ATGAACCTTTTAAAGTGCAGCATTCGCGAAGTGTGTTCTTCGCGAGTCTGCTGCT 60

1 ATGAACCTTTTAAAGTGCAGCATTCGCGAAGTGTGTTCTTCGCGAGTCTGCTGCT 60
 61 GGCCTGTTTCAACATGAGGCGGCGGCGGTATCATTAACGCGCGCATATGTTCCGGC 120
 61 GGCCTGTTTCAACATGAGGCGGCGGCGGTATCATTAACGCGCGCATATGTTCCGGC 120
 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGTTCCGCTTAACGCTGCTGCTGCAA 180
 121 CCGGACTCAACGTTGAGCATTTATCAGTAGAGTTCCGCTTAACGCTGCTGCTGCAA 180
 181 AGCGATCCCGTAAATATGATCAGTGTATCCGCTGTTATCCATGAATAGGACAT 240
 181 AGCGATCCCGTAAATATGATCAGTGTATCCGCTGTTATCCATGAATAGGACAT 240
 241 GCAGCGCAGGTTGCGGATTAATAGTAACTGTAAGTCAAGTGTGTTCAAGAAATAT 300
 241 GTAGGCGCAGGTTGCGGATTAATAGTAACTGTAAGTCAAGTGTGTTCAAGAAATAT 300
 301 GCCACCATGACCATGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAGGTAAG 360
 301 GCCACCATGATAGTGTGTTACCGGTGTTTACCATGAATGACATGACAGCGGT 360
 361 AATTAACCGCGCGGTGTTATCAGACCGCATGATTCAGCGTATGCGTCAAGTT 420
 361 AATTAACCGCGCGGTGTTATCAGACCGCATGATTCAGCGTATGCGTCAAGTT 420
 421 GGTGTTGCAACAGCGCGGTATTAACGATTA 456
 421 GGTGTTGCAACAGCGCGGTATTAACGATTA 456

RESULT 14

AAQ73066 standard; DNA; 361 BP.

AAQ73066;

27-AUG-2003 (revised)

25-MAR-2003 (revised)

26-JUN-1995 (first entry)

Agfa sequence.

Salmonella; Agfa; vaccine; genetic immunization; ds.

Salmonella enteritidis.

Location/Qualifiers

1..359

/*tag= "a"

/*tag= "Agfa"

/*tag= "d"

/*tag= "TAFF5 primer (pair with TAFF6)"

/*tag= "TAFF3 primer (pair with TAFF4)"

/*tag= "TAFF6 primer (pair with TAFF5)"

/*tag= "TAFF4 primer (pair with TAFF3)"

/*tag= "TAFF4 primer (pair with TAFF3)"

/*tag= "TAFF4 primer (pair with TAFF3)"

/*tag= "TAFF4 primer (pair with TAFF3)"

Thu Mar 18 12:28:02 2004

Job time : 246.684 secs

us-09-543-407-19.rng

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:56:17 ; Search time 47.6647 Seconds

(without alignments)
5309.115 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttcaaaagtcg.....ccacgcctaacacgtrataa 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCCTS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	412.8	90.5	456	1	US-08-233-788A-58
2	288.6	63.3	361	1	US-08-233-788A-56
3	34	7.5	1344	3	US-09-120-927-1
4	34	7.5	1344	4	US-09-431-614-5
5	32.8	7.2	1218	4	US-09-489-039A-403
6	32.2	7.1	1399	4	US-09-621-976-8976
7	32	7.0	1152	4	US-09-489-039A-2755
8	31.4	6.9	534	4	US-09-489-039A-2195
9	31.4	6.9	1818	4	US-09-489-039A-2162
10	31.2	6.8	549	4	US-09-252-991A-14225
11	31.2	6.8	696	4	US-09-252-991A-14137
12	31	6.8	2052	1	US-08-443-104-5
13	31	6.8	2052	1	US-08-238-130-6
14	31	6.8	2052	1	US-08-442-859-5
15	31	6.8	2052	2	US-08-398-489-5
16	31	6.8	2052	2	US-08-894-772-1
17	31	6.8	2052	2	US-09-207-844-1
18	31	6.8	2052	4	US-09-252-509-1
19	31	6.8	2052	5	PCT-US95-05534-5
20	30.8	6.8	1662	1	US-08-565-386-2
21	30.8	6.8	2547	3	US-08-508-761B-1
22	30.8	6.8	4529	1	US-08-565-386-1
23	30.8	6.8	1664976	4	US-08-916-421B-1
24	30.6	6.7	4403765	3	US-09-103-840A-2
25	30.6	6.7	4411529	3	US-09-103-840A-1
26	30	6.6	2436	4	US-09-540-236-492
27	30	6.6	3895	4	US-08-961-527-201

C	28	30	6.6	65792	4	US-09-596-002-31	Sequence 31, Appl
C	29	29.8	6.5	1008	4	US-09-252-991A-10904	Sequence 10904, A
C	30	29.8	6.5	1077	4	US-09-252-991A-10818	Sequence 10818, A
C	31	29.8	6.5	1521	4	US-09-252-991A-10865	Sequence 10865, A
C	32	29.8	6.5	51259	3	US-08-781-891A-209	Sequence 209, App
C	33	29.8	6.5	51259	4	US-09-618-166-209	Sequence 209, App
C	34	29.6	6.5	972	4	US-09-328-352-2055	Sequence 2055, Ap
C	35	29.4	6.4	912	4	US-09-543-681A-2040	Sequence 2040, Ap
C	36	29.4	6.4	1252	4	US-09-561-756-29	Sequence 29, Appl
C	37	29.4	6.4	1252	4	US-09-227-721-29	Sequence 29, Appl
C	38	29.4	6.4	1252	4	US-09-954-697-29	Sequence 29, Appl
C	39	29.4	6.4	536165	4	US-09-214-808-1	Sequence 1, Appl
C	40	29.2	6.4	708	4	US-09-543-681A-2551	Sequence 2551, Ap
C	41	29.2	6.4	4403765	3	US-09-103-840A-2	Sequence 2, Appl
C	42	29.2	6.4	4411529	3	US-09-103-840A-1	Sequence 1, Appl
C	43	29	6.4	939	4	US-09-976-594-108	Sequence 108, App
C	44	29	6.4	1097	4	US-09-988-784-7	Sequence 7, Appl
C	45	29	6.4	98844	4	US-09-791-211-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1
US-08-233-788A-58
Sequence 58, Application US/08233788A
Patent No. 5635617
GENERAL INFORMATION:
APPLICANT: Dorian, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
TITLE OF INVENTION: OF SALMONELLA
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 58:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-788A-58
Query Match 90.5%; Score 412.8; DB 1; Length 456;
Best Local Similarity 94.1%; Pred. No. 1.7e-137;
Matches 429; Conservative 0; Mismatches 27; Indels 0;

QY 1 ATGAACCTTTAAAGTGACAGATTGCGACATGTAAGTTCTGCGAGTCTGCTGCT 60
DB 1 ATGAACCTTTAAAGTGACAGATTGCGACATGTAAGTTCTGCGAGTCTGCTGCT 60
QY 61 GCGCGTCTTCCACATAGGGGGGGGGGGGTTATCATATACGCGCGGCAATAGTTCCGGC 120
DB 61 GCGCGTCTTCCACATAGGGGGGGGGGGGTTATCATATACGCGCGGCAATAGTTCCGGC 120
QY 121 CCGACCTCAACGTTGAGCATTTATCAATGACGTTCCGCTAACGCGCGGCTTCTGCAA 180
DB 121 CCGACCTCAACGTTGAGCATTTATCAATGACGTTCCGCTAACGCGCGGCTTCTGCAA 180
QY 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGACAT 240
DB 181 AGCGATCCCGTAAATATGATCAGCTGTTACCCGTTGTTACCATGAATGACAT 240
QY 241 GCGAGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
DB 241 GCGAGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 300
QY 301 GCCACATCGACAGTGAAGAGCTTAAATCTCCGATTAATGATTAATGATTAATGATTAAT 360
DB 301 GCCACATCGACAGTGAAGAGCTTAAATCTCCGATTAATGATTAATGATTAATGATTAAT 360
QY 361 AATAAGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
DB 361 AATAAGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 420
QY 421 GGTTTTGGCAACAGCGGCTTAACAGATTAAT 456
DB 421 GGTTTTGGCAACAGCGGCTTAACAGATTAAT 456

RESULT 2

US-08-233-788A-56
Sequence 56, Application US/08233788A
Patent No. 3635617

GENERAL INFORMATION:

APPLICANT: Doran, James L.
APPLICANT: Kay, William W.
APPLICANT: Collinson, Karen S.
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTION
NUMBER OF SEQUENCES: 61
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233, 788A
FILING DATE: 26-APR-1994
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043, 403C2
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs

TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-788A-56

Query Match
Best Local Similarity 91.3%; Score 288.6; DB 1; Length 361;
Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 64 GTCTTCCACATAGGGGGGGGGGGGTTATCATATACGCGCGGCAATAGTTCCGGCGC 123
DB 1 GTCTTCCACATAGGGGGGGGGGGGTTATCATATACGCGCGGCAATAGTTCCGGCGC 123
QY 124 GACTCAACGTTGACATTTATCAATGACGTTCCGCTAACGCGGCTTCTGCAAAGC 183
DB 61 GACTCAACGTTGACATTTATCAATGACGTTCCGCTAACGCGGCTTCTGCAAAGC 120
QY 184 GATGCCCGTAAATATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 243
DB 121 GATGCCCGTAAATATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 180
QY 244 GCGCAGCGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 303
DB 181 GCGCAGCGTGGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 240
QY 304 ACCATCGACAGTGAAGAGCTTAAATCTCCGATTAATGATTAATGATTAATGATTAAT 363
DB 241 ACCATCGACAGTGAAGAGCTTAAATCTCCGATTAATGATTAATGATTAATGATTAAT 300
QY 364 AACGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 398
DB 301 AACGCGCGGCTCGGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 335

RESULT 3

US-09-120-927-1/c
Sequence 1, Application US/09120927
Patent No. 6262018

GENERAL INFORMATION:

APPLICANT: Kim, Jihyun Francis
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: P.O. Box 1051, Clinton Square
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/120,927
FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION:
APPLICATION NUMBER: US 60/055,108
FILING DATE: 06-AUG-1977
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/1581
TELEPHONE: (716) 263-1304

TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1344 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-09-120-927-1

Query Match 7.5%; Score 34; DB 3; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.09;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGGCGTGTTCACAAATGGGCGCGCGGTAATCATTAACGGCGCGCAAT 111
DB 209 GCTCGGTTGCGCGATTACCTGATGTCGCGATACAGACGATTAAACGTTCCGCCAAT 150
QY 112 AGTTCCGCGCGCGACGATTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
DB 149 AATTGACCATTTGCTCAATGTTGCGCATGATGCGTGTGCGCCACAGCAAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 4
US-09-431-614-5/c
Sequence 5, Application US/09431614
Patent No. 6624139
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Schading, Richard L.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR-INDUCED STRESS
FILE REFERENCE: 21829/41 (EBC-003)
CURRENT APPLICATION NUMBER: US/09/431,614
EARLIER FILING DATE: 1999-11-02
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 5
LENGTH: 1344
TYPE: DNA
ORGANISM: Erwinia amylovora
US-09-431-614-5

Query Match 7.5%; Score 34; DB 4; Length 1344;
Best Local Similarity 54.9%; Pred. No. 0.09;
Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTCTGCTGGCGTGTTCACAAATGGGCGCGCGGTAATCATTAACGGCGCGCAAT 111
DB 209 GCTCGGTTGCGCGATTACCTGATGTCGCGATACAGACGATTAAACGTTCCGCCAAT 150
QY 112 AGTTCCGCGCGCGACGATTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTT 171
DB 149 AATTGACCATTTGCTCAATGTTGCGCATGATGCGTGTGCGCCACAGCAAAATT 90
QY 172 GC 173
DB 89 GC 88

RESULT 5
US-09-489-039A-403
Sequence 403, Application US/09489039A
Patent No. 6610836
GENERAL INFORMATION:
APPLICANT: Gary Breton et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 2709,2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
PRIOR FILING DATE: 1999-01-29
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 403
LENGTH: 1218
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-403

Query Match 7.2%; Score 32.8; DB 4; Length 1218;
Best Local Similarity 58.0%; Pred. No. 0.23;
Matches 58; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 174 TCTGCAAGCGATCCCGTAATATGATCAGCTGTTACCGGTGTTTACCAATGAAT 233
DB 614 TCGTACAGCGCTGATCTCTCAATATGTTACAGACCTTATGCTGTGCGTAATTAAC 673
QY 234 GCACATGACAGCGCGAGGTGCGGATATATGACTATTGAA 273
DB 674 CGACGCGCGCGGACGAGCAGCAAGAAATTAATCGAGAA 713

RESULT 6
US-09-621-976-8976/c
Sequence 8976, Application US/09621976
Patent No. 6639063
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, J.B.
APPLICANT: Jodert, S.
TITLE OF INVENTION: ESTs and Encoded Human Proteins.
FILE REFERENCE: GENSET.054PR2
CURRENT APPLICATION NUMBER: US/09/621,976
EARLIER FILING DATE: 2000-07-21
NUMBER OF SEQ ID NOS: 19335
SOFTWARE: Patent .pm
SEQ ID NO 8976
LENGTH: 399
TYPE: DNA
ORGANISM: Homo sapiens
US-09-621-976-8976

Query Match 7.1%; Score 32.2; DB 4; Length 399;
Best Local Similarity 10.0%; Pred. No. 0.19;
Matches 34; Conservative 151; Mismatches 154; Indels 0; Gaps 0;

QY 88 GGTATCATTAACGGCGCGGCAATAGTTCGCGCGGACCTCAACGTTGACATTTATCAG 147
DB 340 GMYTMYMTSRGSGYRKYTSAMMGRAMKMKKTKMYMKGGKXGSGTYMMRSG 281
QY 148 TAGGTTCCGTAACGGCGGCTGCTGCAAGGCAATGCGGTAATATGATCAGCTG 207
DB 280 STGRWSTRRAMWRSKSGSYRRMAGYRSRMSYSAMWRKMKMTGWSGSR 221
QY 208 GTTACCGGTGTTTACCATGAATGACACATCAGCGCGGCGGTAATATGACT 267
DB 220 TGYAAMWYKSKWCSKSRKMYKKRKRKCTSTKTCYVGSYYKCMKAYTTKKRKY 161
QY 268 ATTGAAGTACTGAAATGTTTCAAGAAATATCCACATGACCAAGTGAACGCTTAA 327
DB 160 RTYYVYKSYMSMKTKRMKTAAYTWTKMKWTCTCMKCTTYMAGTMYRFR 101
QY 328 AACTCCGATATTAATGTCGCGCATATAGCGGTAATAACGCGGCGGCTTAATCAAGCC 387
DB 100 YWYAKAKWSKRCSTWTCYCMKYMAKCWSYMSMKMGKSGSMWKTYYYYYYMK 41
QY 388 GCATCTGATTCACAGCTAATGTCGTCAGGTTGTTT 426
DB 40 WSKTYWMSWMSVARKCMRTYAKTYTWTMTWTKRKYK 2

Query Match	7.0%;	Score 32;	DB 4;	Length 1152;
Best Local Similarity	56.7%;	Pred. No. 0.43;		
Matches	59;	Conservative	0.	Mismatch 11.33;

RESULT 8
US-09-489-039A-2195/c
; Sequence 2195, Application US/09489039A
; Patent No. 6610836

Query Match	6.9%	Score 31.4;	DB 4;	Length 534;
Best Local Similarity	52.7%;	Pred. No. 0.44;		
Matches	68;	Conservative		

RESULT 5

INVENTOR: Gary Bregon et. al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
TITLE OF INVENTION: PNEUMONIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709, 2004/0001
CURRENT APPLICATION NUMBER: 2004/0001

Query Match	6.9%	Score 31.4;	DB 4;	length 1818,
Best Local Similarity	52.7%;	Pred. No. 0.92;		
Matches	68;	Conservative	0;	Mismatches

RESULT 10
US-09-252-991A-14225/c
; Sequence 14225, Application US/09252991A
; Patent No. 6551795

```

; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14225

```

	Query Match	Best Local Similarity	Score 31.2;	DB 4;	Length 549;
	Matches	66; Conservative	0;	Mismatches 58;	Indels 0;
				Gaps	0;
QY	131	CGTTGACATTATCAGTACGATGCCCTAACGCTGGCGCTTCTGCAAAACGATGCC	190		
Db	282	CGTTGACACCGTCAGCGCTGGGGTACGCCGCGATCTCAGTCGATCTGGATATTGGCGCTG	223		
QY	191	GTAATATGATCAGCTGTTTACCCGTTGTTTACCATGAATGGACATGAGGCCACGG	250		
Db	222	GTGCGGACCGTCTGGCGCTGACACGATGGTATGATGAGTGCATGCTCCCAAGTCCAGG	163		

QY 251 GTGC 254
Db 162 GTTC 159

RESULT 11

US-09-252-991A-14137/C
Sequence 14137, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rudenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 14137
LENGTH: 696
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14137

Query Match 6.8%; Score 31.2; DB 4; Length 696;
Best Local Similarity 53.2%; Pred. No. 0.61;
Matches 66; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 131 CGTTGACGATTATTCAGTACGCTTCCCTTAACGCTGCTGCTGTCGAAAGCATGCC 190
Db 377 CGTTGACGCTGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 318
QY 191 GTAATATGATCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 250
Db 317 GTGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 258
QY 251 GTGC 254
Db 257 GTTC 254

RESULT 12

US-08-443-104-5/C
Sequence 5, Application US/08443104
Patent No. 5691162
GENERAL INFORMATION:
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fugleang, Claus
APPLICANT: Branner, Sven
TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5691162 No. 5691162disk of No. 5691162th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/443,104
FILING DATE: 17-MAY-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489
FILING DATE: 03-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Agis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-9655
TELEFAX: 212-867-0123
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-443-104-5

Query Match 6.8%; Score 31; DB 1; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTTATCCCAAGAAATGACATGACGCGGCTGCGGATATATGATCTTGAAC 274
Db 648 GTGAGCTTACGCTGCGGCTGCTGACGCTTACCTTGTATCAGGAGGTATGCTTC 589
QY 275 TGACTCAAAATGTTTCAAAATATGACCATGACAGTGAAGCTTAAACTCCG 334
Db 588 TTGCTCTTCTTACGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 529
QY 335 ATATTAATGTCGCGCAATACGCGGCTGTAATAAGCGCGCTGTTATCAAGCGCATCTG 394
Db 528 GTGCTCTTGAAGAGGCTGACGCGGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 469
QY 395 ATTCCAGCTTAATGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
Db 468 TTGCTCTTGAAGAGGCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 410

RESULT 13

US-08-238-130-6/C
Sequence 6, Application US/08238130
Patent No. 5702934
GENERAL INFORMATION:
APPLICANT: Hastrup, Sven
APPLICANT: Branner, Sven
APPLICANT: Jorgensen, Birthe R.
APPLICANT: Christensen, Tove
APPLICANT: Jorgensen, Birgitte B.
APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
TITLE OF INVENTION: PROCESSES FOR PRODUCING AN ENZYME
TITLE OF INVENTION: ACTIVITY
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: No. 5702934 No. 5702934disk of No. 5702934th America, Inc.
STREET: 405 Lexington Avenue, 64th floor
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10174-6401
COMPUTER READABLE FORM:
MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/238,130
FILING DATE: 04-MAY-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 522/93
FILING DATE: 05-MAY-1993
ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 3965, 200-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: join(1..363, 416..802, 856..1821, 1870..2052)
US-08-238-130-6

Query Match Best Local Similarity 6.8%; Score 31; DB 1; Length 2052;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTGTTACCCATGAAATGACATGACGAGCGGATGATATGACTATTGTAAC 274
DB 648 GTGAGCTTACCTCGGGCTTGTGACGTAACCTTGTGTAACGTAAGGTGAGGTTTC 589
QY 275 TGACTGAGATGTTTCAGAAATATGACCATGACCATGTAAGCTTAAACTCCG 334
DB 588 TTGCTCTTCTTTCAGCCTTGCGCTTGTGACCTCAAGGGGAGAAAGAACTCAACG 529
QY 335 ATATTACTGTGCGCAATAGCGGGTAATAGCGCGCTGTTATAGACCGCATCTG 394
DB 528 GTGTCTTGAAGCGTCGACGGGGTCTTCTGTCACCTGTGTAAGAGACCGGAATC 469
QY 395 ATTCCAGGTAATGTTGCGTCAAGTGTGTTTGGCAACAGCCGCGCTAACCATGAT 453
DB 468 TTGCCCTGTGAAAGCTGTTCCGTAGGAGAAAGCTCGCGCTGACGCCAATCTGTAT 410

RESULT 14

US-08-442-859-5/C
Sequence 5, Application US/08442859
Patent No. 5807729

GENERAL INFORMATION:

APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Braner, Sven

TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 58077290 No. 5807729disk of No. 5807729th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/442,859

FILING DATE: 17-MAY-1995

CLASSIFICATION: 252

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/398,489

FILING DATE: 03-MAR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.

REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180, 010-US

TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655

INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA

US-08-442-859-5

Query Match Best Local Similarity 6.8%; Score 31; DB 1; Length 2052;
Matches 109; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 215 GTGTTGTTACCCATGAAATGACATGACGAGCGGATGATATGACTATTGTAAC 274
DB 648 GTGAGCTTACCTCGGGCTTGTGACGTAACCTTGTGTAACGTAAGGTGAGGTTTC 589
QY 275 TGACTGAGATGTTTCAGAAATATGACCATGACCATGTAAGCTTAAACTCCG 334
DB 588 TTGCTCTTCTTTCAGCCTTGCGCTTGTGACCTCAAGGGGAGAAAGAACTCAACG 529
QY 335 ATATTACTGTGCGCAATAGCGGGTAATAGCGCGCTGTTATAGACCGCATCTG 394
DB 528 GTGTCTTGAAGCGTCGACGGGGTCTTCTGTCACCTGTGTAAGAGACCGGAATC 469
QY 395 ATTCCAGGTAATGTTGCGTCAAGTGTGTTTGGCAACAGCCGCGCTAACCATGAT 453
DB 468 TTGCCCTGTGAAAGCTGTTCCGTAGGAGAAAGCTCGCGCTGACGCCAATCTGTAT 410

RESULT 15

US-08-398-489-5/C
Sequence 5, Application US/08398489
Patent No. 5843753

GENERAL INFORMATION:

APPLICANT: Shuster, Jeffrey R.
APPLICANT: Madden, Mark
APPLICANT: Moyer, Donna L.
APPLICANT: Fuglsang, Claus
APPLICANT: Braner, Sven

TITLE OF INVENTION: NOVEL METALLOPROTEASE HAVING INCREASED
NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 58437530 No. 5843753disk of No. 5843753th America, Inc.
STREET: 405 Lexington Avenue, 64th Floor
CITY: New York
STATE: New York
COUNTRY: USA

ZIP: 10174-6401

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/398,489

FILING DATE: 03-MAR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Agtis Dr., Cheryl H.

REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 4180.010-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2052 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-398-489-5

Query Match 6.8%; Score 31; DB 2; Length 2052;
Best Local Similarity 45.6%; Pred. No. 1.4; Mismatches 130; Indels 0; Gaps 0;
Matches 109; Conservative 0;

Oy	215	GTGTTGTTACCATGAATGCGACATGCGAGGCGGATATAGTACTATTGAAC	274
Db	648	GTGAGCTTAGCTCGGGCTTGCTGACGCTTGTGTAACGTAAGTGTAGTGTTC	589
Oy	275	TGACTCAGATGTTTCAGAAATAATGCCACCATGACCAATGAAACGCTAAAACCTCG	334
Db	588	TGCTCTTCTTCTCAGCCTTGCCCTTGTCAAGCTCAACGGGAGAGAAAGAACATCAACG	529
Oy	335	ATATTACTGTGGCCCAATACGGCGGTAAATACGCCCGCTGTTATCAGACCGCATCTG	394
Db	528	GTGTCCTTGAGAGCGTCGACGGGGCTTTCGTGCACGCTGTGTAAGAGAACCGGGAATC	469
Oy	395	ATTCCAGCGTATGTTGCGTCAGTTGTTTGGCAACAGCCCAAGGCTAACCGATAT	453
Db	468	TTGCCCTGTGAAAGCTGTTTCGTAGAGAAAGACCTGCGCGTCAGCGCAATCTGTAT	410

Search completed: March 16, 2004, 04:37:54
Job time: 58.6647 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 401.636 Seconds
(without alignments)

4179.927 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456
Sequence: 1 atgaacttttaaaagtcgc.....ccacgcgtaccagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2432557 seqs, 1840798884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*

2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*

3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*

4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*

5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*

6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*

7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*

8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*

9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:*

10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:*

11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:*

12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*

13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:*

14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:*

15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:*

16: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*

17: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*

18: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	34.6	7.6	1662	14	US-10-156-761-3483
2	34.6	7.6	9025608	14	US-10-156-761-1
3	34	7.5	1344	9	US-09-835-684-6
4	34	7.5	1344	9	US-09-880-371-6
5	34	7.5	1344	9	US-09-879-248-5
6	34	7.5	1344	14	US-10-010-390-6
7	34	7.5	1344	15	US-10-441-736-5
8	33.4	7.3	655	14	US-10-184-644-418
9	33.4	7.3	655	14	US-10-184-634-418
10	33.2	7.3	1848	14	US-10-156-761-7222
11	33.2	7.3	9025608	14	US-10-156-761-1
12	32.8	7.2	24081	14	US-10-132-134-13
13	32.8	7.2	52101	14	US-10-132-134-1
14	32.2	7.1	1083	15	US-10-369-493-33133
15	31.8	7.0	930	9	US-09-815-242-6028

16	31.8	7.0	930	12	US-10-282-122A-20335	Sequence 20335, A
17	31.8	7.0	930	15	US-10-369-493-24480	Sequence 24480, A
18	31.8	7.0	978	15	US-10-369-493-24265	Sequence 24265, A
19	31.6	6.9	6003	13	US-10-010-901-9	Sequence 9, Appl1
20	31.4	6.9	460	15	US-10-260-238-787	Sequence 787, App
21	31.4	6.9	7374	10	US-09-764-891-10207	Sequence 10207, A
22	31.4	6.9	7374	14	US-10-205-428-1006	Sequence 1006, Ap
23	31.2	6.8	669	14	US-10-156-761-687	Sequence 687, App
24	31.2	6.8	2266	15	US-10-108-260A-375	Sequence 375, App
25	31	6.8	1049	14	US-10-123-155-358	Sequence 358, App
26	31	6.8	1049	14	US-10-146-731-358	Sequence 358, App
27	31	6.8	1049	14	US-10-140-472-358	Sequence 358, App
28	31	6.8	1049	14	US-10-141-761-358	Sequence 358, App
29	31	6.8	1049	14	US-10-142-885-358	Sequence 358, App
30	31	6.8	1049	14	US-10-158-790-358	Sequence 358, App
31	31	6.8	1049	15	US-10-137-871-358	Sequence 358, App
32	31	6.8	1049	15	US-10-140-923-358	Sequence 358, App
33	31	6.8	1049	15	US-10-141-756-358	Sequence 358, App
34	31	6.8	1049	15	US-10-141-759-358	Sequence 358, App
35	31	6.8	1049	15	US-10-140-805-358	Sequence 358, App
36	31	6.8	1049	15	US-10-140-864-358	Sequence 358, App
37	31	6.8	1841	15	US-10-369-493-45479	Sequence 45479, A
38	30.8	6.8	524	14	US-10-029-386-4400	Sequence 4400, A
39	30.8	6.8	715	12	US-10-424-599-10132	Sequence 10132, A
40	30.8	6.8	1575	10	US-09-873-367C-255	Sequence 255, App
41	30.6	6.7	1164	12	US-10-282-122A-13500	Sequence 13500, A
42	30.6	6.7	1421	12	US-10-425-114-12527	Sequence 12527, A
43	30.6	6.7	1573	9	US-09-771-161A-1	Sequence 1, Appl1
44	30.6	6.7	1668	15	US-10-355-956-3	Sequence 3, Appl1
45	30.6	6.7	75216	14	US-10-080-170-646	Sequence 646, App

ALIGNMENTS

RESULT 1

US-10-156-761-3483/C

Sequence 3483, Application US/10156761

Publication No. US20030119018A1

GENERAL INFORMATION:

APPLICANT: OMDRA, SATOSHI

APPLICANT: IKEDA, HARUO

APPLICANT: ISHIKAWA, JUN

APPLICANT: HORIKAWA, HIROSHI

APPLICANT: SHIBA, TADAYOSHI

APPLICANT: SAKAKI, YOSHIYUKI

APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES

FILE REFERENCE: 249-262

CURRENT APPLICATION NUMBER: US/10/156,761

CURRENT FILING DATE: 2002-05-29

PRIOR APPLICATION NUMBER: JP 2001-204089

PRIOR FILING DATE: 2001-05-30

PRIOR APPLICATION NUMBER: JP 2001-272697

PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109

SEQ ID NO 3483

LENGTH: 1662

TYPE: DNA

ORGANISM: Streptomyces avermitilis

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(1662)

US-10-156-761-3483

Query Match 7.6%; Score 34.6; DB 14; Length 1662;

Best Local Similarity 49.7%; Pred. No. 0.78; Mismatches 89; Indels 0; Gaps 0;

Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Oy 50 GTGCTTGGCTGGCGTCTTCCACATGCGCGCGGCTATCATACGCGCGGCA 109

Db 590 GCGCGCGGTGACGTCGATACAGATCGCGCGCGCTTCATCGTCACGCGCGGCGCG 531

QY 110 ATAGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGTAACGCTGGCC 169
 DB 530 TGGGGGCGCCCGCCAGTCCCGGAGTCCCGGAGTGTATCTGTCCGCGAGGGTGGCG 471
 QY 170 TTGCTTCGCAAGCGATGCGCGTAATAATATGATCAGCTGGTTACCGGTGTTTACC 226
 DB 470 TTGAACCTTCCTGGGGGAGCGGCGGAGACGTCCTGACGTCCTCGGAGGATGCC 414

RESULT 2

US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; PRIOR FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermectilis
 ; FEATURE:
 ; NAME/KEY: misc-feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Query Match
 Best Local Similarity 7.6%; Score 34.6; DB 14; Length 9025608;
 Matches 88; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 50 GTGCTGTGCTGCGCTGCTTCACATGAGGGGCGGGGTAATCATACGCGCGCGCA 109
 DB 4331781 GCGCGCGGGGTGCGATGATGCGGACGCGGTCTTCATGTCACGCGCGCGCGC 4331840
 QY 110 ATAGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGTAACGCTGGCC 169
 DB 4331841 TGGGGGCGCCCGCCAGTCCCGGAGTCCCGGAGTGTATCTGTCCGCGAGGGTGGCG 4331900
 QY 170 TTGCTTCGCAAGCGATGCGCGTAATAATATGATCAGCTGGTTACCGGTGTTTACC 226
 DB 4331901 TTGAACCTTCCTGGGGGAGCGGCGGAGACGTCCTGACGTCCTCGGAGGATGCC 4331957

RESULT 3

US-09-835-684-6/c
 ; Sequence 6, Application US/09835684
 ; Patent No. US20020019337A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Qiu, Dewen
 ; APPLICANT: Remick, Dean
 ; TITLE OF INVENTION: TREATMENT OF FRUITS OR VEGETABLES WITH HYPERSENSITIVE
 ; TITLE OF INVENTION: RESPONSE ELICITOR TO CONTROL POSTHARVEST DISEASE OR
 ; FILE REFERENCE: 21829/71
 ; CURRENT APPLICATION NUMBER: US/09/835,684
 ; PRIOR FILING DATE: 2001-04-16
 ; PRIOR APPLICATION NUMBER: 60/198,359
 ; PRIOR FILING DATE: 2000-04-19

NUMBER OF SEQ ID NOS: 12
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-835-684-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGCGTGGCGCTGCTTCACATGAGGGGCGGGGTAATCATACGCGCGCAAT 111
 DB 209 GCTCGGTTGGCGCATTAACGATGTCGAGATAGAGATTAACGTTTCCGCAAT 150
 QY 112 AGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGTAACGCTGGCCTT 171
 DB 149 AATTGACCATTTGCTCAATGATGTTGCCGATGATGATGTTGCCCAACCGCAATTT 90
 QY 172 GC 173
 DB 89 GC 88

RESULT 4

US-09-880-371-6/c
 ; Sequence 6, Application US/09880371
 ; Patent No. US20020059658A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Wei, Zhong-Min
 ; APPLICANT: Derocher, Jay
 ; TITLE OF INVENTION: METHODS OF IMPROVING THE EFFECTIVENESS OF TRANSGENIC
 ; FILE REFERENCE: 21829/91
 ; CURRENT APPLICATION NUMBER: US/09/880,371
 ; PRIOR FILING DATE: 2001-06-13
 ; PRIOR APPLICATION NUMBER: 60/211,585
 ; PRIOR FILING DATE: 2000-06-15
 ; NUMBER OF SEQ ID NOS: 16
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1344
 ; TYPE: DNA
 ; ORGANISM: Erwinia amylovora
 US-09-880-371-6

Query Match
 Best Local Similarity 7.5%; Score 34; DB 9; Length 1344;
 Matches 67; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 52 GCTTGCGTGGCGCTGCTTCACATGAGGGGCGGGGTAATCATACGCGCGCAAT 111
 DB 209 GCTCGGTTGGCGCATTAACGATGTCGAGATAGAGATTAACGTTTCCGCAAT 150
 QY 112 AGTTCGGCCCGGACTCAAGTTGACATTTATCAGTTCGGTTCGTAACGCTGGCCTT 171
 DB 149 AATTGACCATTTGCTCAATGATGTTGCCGATGATGATGTTGCCCAACCGCAATTT 90
 QY 172 GC 173
 DB 89 GC 88

RESULT 5

US-09-879-248-5/c
 ; Sequence 5, Application US/09879248
 ; Patent No. US20020062500A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Fan, Hao
 ; APPLICANT: Wei, Zhong-Min
 ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITING DOMAINS AND USE
 ; TITLE OF INVENTION: THEREOF

Query Match	7.3%	Score 33.4	DB 14	Length 655
Best Local Similarity	14.9%	Pred. No. 1.2		
Matches 39	Conservative 76	Mismatches 147	Indels 0	Gaps 0
QY	18	GGAGCAGTTGGCAGCAATGCTAGTTTGGCAGTGCCTCGAGCCTCGTTCCACATG	77	
	:::	: : ::::	: : : : :	: : : : :

```

/ Sequence 418, Application US/10184634
/ Publication No. US20030068684A1
/ GENERAL INFORMATION:
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Dian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L.
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K.
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE OF INVENTION: ACIDS ENCODING THE SAME
/ FILE REFERENCE: P3440R1C217
/ CURRENT APPLICATION NUMBER: US/10/184,634
/ Prior Application removed - See File Wrapper or Palm
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO. 418
/ LENGTH: 655
/ TYPE: prt
/ ORGANISM: Homo Sapien
US-10-184-634-418

```

QY 18 GCGAGCAATTCGCACGAATCTGTAAGTTTCTTGCGAGTGTCTTGCGTGGCGTGTCCGATCACAANTG 77
Db : : : : : Indels 0; Gaps 0;
325 SSM .CCAT .KS.. M. DYPD.C. IDNBH.N. AMNBH.MCYTSSANH.BAC...SS.B 266
QY GGGCGGGGGGGGATCATCAATCAAGCGGGGGGGGCAATAGTTCCGCCCGGACTCAAGCTTAG 137
Db M.YESTISSN.ABNKCM.BRASS.B.DA.KD.X...TAC.S.SASS.S.ACGBNHA. 206
QY CATTATACAGTACCGTTCCGCTAAAGCGTGGCCGCTGTCCGAAAAGCATGCCGCTTAATA 197
Db MAC.M.BB.N.S.GHARTGMKBSS.BHS.ACTTG.MGYBH.A.AOMVBCWB.GRAD. 146
QY TGATCAGCTGTTACCCGCTGTTTGATCCCATGAAATGGCAATGCAAGGCCAGGGTGGCGA 257
Db TGTATNS..KC..GAG.YHA.TTG..M..K.W.G..SGHDGM..CW.DYA.ACB.GSSGBY 86
QY 258 TATATGACTACTGAATGACTGACT 279
Db : : : : :
85 .SANAGD.SBRAC.T.GMHGA.B 64

```

? Sequence 7222, Application us/10156761
? Publication No. US20030119018A1
? GENERAL INFORMATION:
? APPLICANT: OMDA, SATOSHI
? APPLICANT: IKEDA, HARUO
? APPLICANT: ISHIKAWA, JUN
? APPLICANT: HORIKAWA, HIROSHI
? APPLICANT: SHIBA, TADAYOSHI
? APPLICANT: SAKAKI, YOSHIHIKI
? APPLICANT: HATTORI, MASAHIRA
? TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
? FILE REFERENCE: 249-262
? CURRENT APPLICATION NUMBER: US/10156, 761
? CURRENT FILING DATE: 2002-05-29
? PRIOR APPLICATION NUMBER: JP 2001-204089
? PRIOR FILING DATE: 2001-05-30
? PRIOR APPLICATION NUMBER: JP 2001-272697
? PRIOR FILING DATE: 2001-09-02
? NUMBER OF SEQ ID NOS: 15109
? SEQ ID NO 7222
? LENGTH: 1848
? TYPE: DNA
? ORGANISM: Streptomyces avermitilis
? FEATURES:
? NAME/KEY: CDS
? LOCATION: (1)..(1848)
US-10-156-761-7222

```

	96; Conservative	0; Mismatches	108; Indels	0; Gaps
QY	46	GGCAGTCTGCTGCTGGCCGCTGTTCCAACTAGGGCGCGCGGCTATCATTAAGCGCGC	105	
Db	834	GGCGGGTACACCGCGGACAGCGAGGCCCGCGGTCCGGAGCCGACATAGAGACGTGCGC		
QY	106	GGCAATAGTTCGGGCCCGGACTCAACGTTATGACATTTATAGTACGTTCCGCTAACGT	165	
Db	774	GTACCGGTTTCCGGCGGGGAGTGAACGACGCCGATAGCCGGGCCAAGTTCGCGTTGCT		
QY	166	GCCTTGCCTCTSCAAGAAGCATGCCCGTAAATATGATACGCTGGTTAACCGGTGTTTACC	225	
Db	714	GGGTTTGCGCAACAGGGGTGTGTCTCCGCGCAGATGACGACCGGTGTCTCTCCGGGCGCAC	655	
QY	226	CATGAATGGACATGACGCGCAGGG	251	
Db	654	GCCTGAACGGCGCAGCAGGTCTTCGG	629	

1 Sequence 1, Application US/10156761
2 Publication No. US20030119018A1
3 GENERAL INFORMATION:
4 APPLICANT: OMEGA, SATOSHI
5 APPLICANT: IKEDA, HARUO
6 APPLICANT: ISHIRAWA, JUN
7 APPLICANT: HORIKAWA, HIROSHI
8 APPLICANT: SHIANT, TADAYOSHI
9 APPLICANT: SHIBA, TADAYOSHI
10 APPLICANT: SAKAKI, YOSHIYUKI
11 APPLICANT: HATTORI, MASAHIRA
12 TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
13 FILE REFERENCE: 249-262
14 CURRENT APPLICATION NUMBER: US/10156,761
15 CURRENT FILING DATE: 2002-05-29
16 PRIOR APPLICATION NUMBER: JP 2001-204089
17 PRIOR FILING DATE: 2001-05-30
18 PRIOR APPLICATION NUMBER: JP 2001-272697
19 PRIOR FILING DATE: 2001-08-02
20 NUMBER OF SEQ ID NOS: 15109
21 SEQ ID NO 1


```

; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6028
; LENGTH: 930
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(930)
US-09-815-242-6028

```

```

Query Match          7.0%; Score 31.8; DB 9; Length 930;
Best Local Similarity 52.7%; Pred. No. 5.2;
Matches 69; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 311 ACCAGTGAACGCTAAACTCCGATATTACTGTGCGCAATACGCGGTATTAACGCCG 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 233 AGCAGCAGCGCGGTAAAGCAGCAGCATGATGCGCGTCAAGCCTGGGGAAATCTCCG 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 371 CGCTGTATATCAGACCGCATCTGATTCAGCGTAATGTCGCTGAGTTGTTGGCA 430
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 293 CGCTGTGTTGGCGTGTGATGATGATTTGCTGATGCGGTGCTGTGATGCGCG 352
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 431 ACAAGCCACG 441
   ||||| |||||
DB 353 GCAAGTTCATG 363
   ||||| |||||

```

Search completed: March 17, 2004, 08:16:43
 Job time : 424.736 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 15, 2004, 16:50:22 ; Search time 2230.91 Seconds
(without alignments) 6103.863 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttttaaaagfsgc.....ccacgctaccagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqe, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estcu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_ges_hum:*
18: em_ges_inv:*
19: em_ges_pln:*
20: em_ges_vit:*
21: em_ges_fun:*
22: em_ges_mam:*
23: em_ges_mus:*
24: em_ges_pro:*
25: em_ges_rod:*
26: em_ges_phg:*
27: em_ges_vit1:*
28: gb_ges1:*
29: gb_ges2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
C 1	237.4	52.1	680 12	BU618688 BU618688
C 2	37.4	8.2	484 9	AV430994 AV430994
C 3	35.4	7.8	664 13	CA141136 CA141136
C 4	34.8	7.6	675 13	BX251172 BX251172

C 5	34.6	7.6	699 12	BI959328 BI959328
C 6	34.2	7.5	296 9	AV058630 AV058630
C 7	34.2	7.5	669 28	AZ959284 AZ959284
C 8	34.2	7.5	978 28	B2569281 B2569281
C 9	34.2	7.5	1036 12	BG671547 DRNBUB05
C 10	34.2	7.5	1137 28	B2565783 B2565783
C 11	34.2	7.5	234 10	BB581987 BB581987
C 12	34	7.5	500 14	CA711477 CA711477
C 13	34	7.5	500 28	A2180696 A2180696
C 14	33.8	7.4	365 9	AA746477 AA746477
C 15	33.8	7.4	558 14	CD374421 CD374421
C 16	33.8	7.4	559 13	BU765156 BU765156
C 17	33.8	7.4	902 28	B2569250 B2569250
C 18	33.8	7.4	941 29	CG178280 CG178280
C 19	33.8	7.4	1098 29	CNS02GDP CNS02GDP
C 20	33.6	7.4	680 10	BF668888 BF668888
C 21	33.6	7.4	966 13	BQ718166 BQ718166
C 22	33.4	7.3	505 6	AL809250 AL809250
C 23	33.4	7.3	523 12	BJ334624 BJ334624
C 24	33.4	7.3	715 13	BU444683 BU444683
C 25	33.4	7.3	1064 13	BQ939900 BQ939900
C 26	33.2	7.3	317 12	BM333324 BM333324
C 27	33.2	7.3	625 10	BF501499 BF501499
C 28	33.2	7.3	629 9	AU295325 AU295325
C 29	33.2	7.3	806 13	BX888120 BX888120
C 30	33	7.2	509 12	BM896625 BM896625
C 31	33	7.2	649 12	BI261742 BI261742
C 32	33	7.2	826 28	B2462870 B2462870
C 33	32.8	7.2	521 14	CA703231 CA703231
C 34	32.8	7.2	572 12	BI66794 BI66794
C 35	32.8	7.2	594 12	BJ335653 BJ335653
C 36	32.8	7.2	613 14	CB935126 CB935126
C 37	32.8	7.2	653 12	BU336964 BU336964
C 38	32.8	7.2	1893 28	AY414592 Pan trogl
C 39	32.6	7.1	204 28	AZ577991 AZ577991
C 40	32.6	7.1	352 9	AA784257 AA784257
C 41	32.6	7.1	356 9	AA785291 AA785291
C 42	32.6	7.1	778 29	CNS03735 CNS03735
C 43	32.6	7.1	813 28	B2204853 B2204853
C 44	32.6	7.1	919 14	CK246169 CK246169
C 45	32.6	7.1	942 14	CK247920 CK247920

ALIGNMENTS

RESULT 1
BU618688/c 680 bp mRNA linear EST 01-OCT-2003
DEFINITION BU618688 NIBB Mochii normalized Xenopus early gastrula library
Xenopus laevis cDNA clone XLA186D22 5', mRNA sequence.
BU618688
BU618688.1 GI:37256713

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

Xenopus laevis (African clawed frog)
Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 680)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Kltayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-I,T. and
Kohara,Y.
Expressed genes in X. laevis embryo
Unpublished (2001)
Contact: Tadasi Shin-I
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6886
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp
The information of this clone is available through the following
URL.

FEATURES

http://xenopus.nibb.ac.jp.
Location/Qualifiers
1. .680
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="X1186D22"
/tissue_type="whole embryo"
/dev_stage="stage 10.5"
/clone_lib="NIBB Mochii normalized Xenopus early gastrula library"

ORIGIN

Query Match 52.1%; Score 237.4; DB 12; Length 680;
Best Local Similarity 70.0%; Pred. No. 6.4e-57;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGGCGACGATTCGACGACATGATGTTCTGGCAGTCTTGGCT 60
DB 593 ATGAACCTTTAAAGGCGACGATTCGACGACATGATGTTCTGGCAGTCTTGGCT 60
QY 61 GGCGCTCTTCAACATAGGCGCGCGCGGTATCATACGGCGCGCATATAGTTCCGCG 534
DB 533 GGCGCTCTTCAACATAGGCGCGCGGTATCATACGGCGCGCATATAGTTCCGCG 534
QY 121 CCGGACTCAACGTTGAGCATTTTATCATGAGTTCGCTTCCGCTTCCGCTTCCGCT 474
DB 473 CCAATTTGAGTGAACATTTACAGTACGCGTGGCGTAACTTTCGCTTCCGCT 474
QY 181 AGCGATCCCGTAAATATGATCATGCTGTTACCGCTGTTGTTTACCATGAAATGCAAT 414
DB 413 ACTGATGCGCCGTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 414
QY 241 GCGGCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 354
DB 353 GTTGGTCAAGGCGTCAAGTGAAGTCAATGATGATGATGATGATGATGATGATGAT 354
QY 301 GCGGCGCGGCGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 294
DB 293 GCTACTCTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
QY 361 AATAACCGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 234
DB 233 GCGGCGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTTCGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 174
DB 173 GCGTTCGCGGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 138

RESULT 2

AV430994/c

LOCUS

484 bp

mRNA

linear EST 23-AUG-2000

DEFINITION

AV430994 Porphyra yezoensis TU-1 Porphyra yezoensis cDNA clone

ACCESSION

AV430994.1 GI:8586219

VERSION

EST

KEYWORDS

Porphyra yezoensis

SOURCE

Porphyra yezoensis

ORGANISM

Porphyra yezoensis

REFERENCE

1 (bases 1 to 484)

AUTHORS

Nikaido, I., Asamizu, E., Nakajima, M., Nakamura, Y., Saga, N. and Tabata, S.

TITLE

Generation of 10,154 expressed sequence tags from a leafy

JOURNAL

DNA Res. 7, 223-227 (2000)

MEDLINE

20363100

PUBMED

10907854

COMMENT

Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute

FEATURES

source

Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL: http://www.kazusa.or.jp/en/plant/.
Location/Qualifiers
1. .484
/organism="Porphyra yezoensis"
/mol_type="mRNA"
/strain="TU-1"
/db_xref="taxon:2788"
/clone="P1027e10.x"
/clone_lib="Porphyra yezoensis TU-1"
/note="Vector: pBluescriptSK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 8.2%; Score 37.4; DB 9; Length 484;
Best Local Similarity 57.1%; Pred. No. 10;
Matches 68; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 321 CGTAAACTCCGATATTAATGTCGCGCAATACGCGGATTAATACCGCGCTGTTAA 380
DB 441 CCAAAACACAAAGTGTCCGCGCTCAAGACGCGCTCAATACGCGCTGCGCAG 382
QY 381 TCAGACCGCATGATTCAGCGTATGTCGTCAGTGTGTTGGCAACAACGCCA 439
DB 381 GCGGAAGCTGTGTATCAAGTATGTCGTCAGGTCGCGGAAGGCGCTCCCGGCA 323

RESULT 3

CA141136/c

LOCUS

664 bp

mRNA

linear EST 24-SEP-2003

DEFINITION

SCUFR2058D10.9 RT2 Saccharum officinarum cDNA clone SCUFR2058D10

ACCESSION

CA141136.1 GI:35034821

VERSION

EST

KEYWORDS

Saccharum officinarum

SOURCE

Saccharum officinarum

ORGANISM

Saccharum officinarum

REFERENCE

1 (bases 1 to 664)

AUTHORS

Vettore, A.L., da Silva, F.R., Kemper, B.L. and Arruda, P.

TITLE

The libraries that made SUCEST

JOURNAL

Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT

Centro de Biologia Molecular e Engenharia Genetica

UNIVERSITY

Universidade Estadual de Campinas

ADDRESS

Caixa Postal 6010, 13083-970, Campinas SP, Brazil

TELEPHONE

Tel: 55 19 3788 1137

FAX

Fax: 55 19 3788 1089

EMAIL

Email: parvada@unicamp.br

CLONE DISTRIBUTION

clone distribution: clone distribution information can be found

HTTP

http://www.bcccenter.itav.unesp.br

PLATE

Plate: 058 Row: D Column: 10

SEQ PRIMER

Seq primer: T7 Promoter Primer.

LOCATION/QUALIFIERS

1. .664

/organism="Saccharum officinarum"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCUFR2058D10"

/lab_host="DH10B"

/note="Organ: Root tips (0.3cm-long) from adult plants; undirectional cDNA library generated from [Root tips (0.3cm-long) from adult plants]. cDNA was prepared (Invitrogen). The double-strand cDNAs were fractionated between 0.8 and 1.5 kb were directionally cloned into the vector. Details of each source of RNA and library

construction can be obtained at
http://sucsec.lad.ic.unictamp.br/public"

Query Match 7.8%; Score 35.4; DB 13; Length 664;
Best Local Similarity 46.2%; Pred. No. 44;
Matches 117; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

QY 22 GCATTCGACGAAATGTAAGTTCTTGGCAGTGTCTGGCTGGGCTGTCCCAATGGGCG 81
DB 521 GCTTCGAGTATGTGGAGCTGAGCACCCTGTTCCGTTGACAGACCTTGTACATGGCCAC 462
QY 82 GCGCGCGGTAAATCATATACGCGCGGCAATAGTTCCGCGCGGCACTCAACGTTGAGCATT 141
DB 461 GTGCGCGCGGCGCGCATCCCGCGCGGCAACCCCGCGCTGGCCAAACCTGAGCCCC 402
QY 142 TATCAGTACGCTTCGCTTACGCTGCGCTTGTCTGCAAGAGATGCCCGTAATATGAT 201
DB 401 CTGACGACGCGCTCGGTGGCGCTGCTGCGGCTGTGCGCGCTGGCGCGCTCATGCAC 342
QY 202 CAGCTGCTTACCCGCTGTCTTATACCATGAAATGACATGACGCGCAGGCTCGCATAT 261
DB 341 CCGCGACGAGAGATTGTCTGCGCGCGGTGGGACGAGCATGAAGAGCGAGCTCCGATGAG 282
QY 262 AGTACTATTGAAC 274
DB 281 CTGTCTTGCAC 269

RESULT 4 675 bp mRNA linear EST 24-FEB-2003
BX251172
LOCUS BX251172 Pinus pinaster differentiating xylem adult Pinus pinaster
DEFINITION CDNA clone P0046C03, mRNA sequence.
ACCESSION BX251172
VERSION BX251172.1 GI:28513831
KEYWORDS EST.
SOURCE Pinus pinaster
ORGANISM Pinus pinaster
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Coniferopsida; Coniferales; Pinaceae; Pinus; Pinus.
1 (bases 1 to 675)
Canton, F.R., Le Provost, G., Garcia, V., Barre, A., Frigerio, J.-M.,
Patva, J., Fevereiro, P., Avila, C., Mouret, J.-F., Brach, J., de
Dauvergne, A., Canovas, F.M., and Plomion, C.
Transcriptome analysis of wood formation in maritime pine
Unpublished (2003)
TITLE Contact: Frigerio JM
JOURNAL Genetic et Amelioration 69
COMMENT INRA
route d'Arcachon 33612 Cestas CEDEX France
Email: frigerio@pierroton.inra.fr
Email: frigerio@pierroton.inra.fr
Location/Qualifiers

FEATURES

1..675
/organism="Pinus pinaster"
/mol_type="mRNA"
/strain="ecotype: Corsican"
/db_xref="taxon:71647"
/clone="P0046C03"
/tissue_type="differentiating xylem"
/dev_stages="adult"
/clone_lib="Pinus pinaster differentiating xylem adult"
/note="Vector: Uni-zap XR lambda (Stratagene); Site 1: Eco
RI; Site 2: Xho I; A composite cDNA library was made with
mRNA isolated from normal, compression, opposite, early
and late wood of Maritime pine uni-directionally cloned
into Uni-zap XR using the ZAP-CDNA Synthesis kit
(Stratagene). pBluescript SK(-) plasmids were obtained by
in vivo mass excision. The nucleotide sequence of the
5' end was obtained by automated sequencing with the T3
primer by GENOME EXPRESS, Meylan, France"

ORIGIN

Query Match 7.6%; Score 34.8; DB 13; Length 675;
Best Local Similarity 57.3%; Pred. No. 66;
Matches 63; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 192 TAAATATGATGAGCTGTTACCGTGTGTTACCATTAATGACATGACAGGCCAGG 251
DB 526 TAAATTCCTAGATAGTATGTTATTTTATGACATGACATGACATGATGCTGCG 585
QY 252 TGGGATTAATGACTATTTGAACTGACTCAGATAGTGTTCGAAATTAATG 301
DB 586 TTGATGATGATGTAGAGTGTACCAACCATATATACCCGACATGATATATG 635

RESULT 5 699 bp mRNA linear EST 22-OCT-2001
B1959328
LOCUS B1959328/c
DEFINITION HVSMEN0019E09f Hordeum vulgare rachis EST library HVCDNA0015
(normal) Hordeum vulgare subsp. vulgare cDNA clone HVSMEN0019E09f,
mRNA sequence.
B1959328
VERSION B1959328.1 GI:16310583
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 699)
Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Chin, A., Begum, D.,
Frisch, D., Atkins, M., Yu, Y., Henry, D., Palmer, M., Rambo, T.,
Simmons, J., Oates, R., and Main, D.
Development of a genetically and physically anchored EST resource
for barley genomics: Morex rachis cDNA library
Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 275
Seq primer: AATTAACCTCAGCAACGCG
High quality sequence stop: 452.
Location/Qualifiers

FEATURES

1..699
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="Morex"
/sub_species="vulgare"
/db_xref="taxon:112509"
/clone="HVSMEN0019E09f"
/tissue_type="Rachis"
/lab_host="TJ121"
/clone_lib="Hordeum vulgare rachis EST library HVCDNA0015
(normal)"
/note="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; Plants were grown at Washington State University,
Pullman, WA in a greenhouse, the rachises were excised and
frozen in liquid nitrogen (Kleinhofs lab). In the TJ Close
lab at the University of California, Riverside total RNA
was prepared, poly(A) was purified, one primary
unamplified cDNA library was made, and 1 million pfu were
in vivo excised to give pBluescript SK(-) cDNA phagemids
(Chin). Phagemids were plated and picked at the Clemson
University Genomics Institute (CUGI) (Begum, Palmer,
Frisch, Atkins and Wing). Plasmid DNA preparations, DNA
sequencing and sequence analysis were performed at CUGI
(Wing, Yu, Frisch, Henry, Simmons, Rambo, Main). The
sequence has been trimmed to remove vector sequence and
contains a minimum of 100 bases of piped value 20 or
above. For more details on library preparation and

```
location/Qualifiers
1. 296
/organism="Mus musculus"
/mol_type="rRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810054H02"
/sex="male"
/tissue_type="pancreas"
/dev_stage="adult"
```

```

/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUCG2M0226G20"
/sex="Female"
/lab_host="E. coli strain XL10-Gold, T1-resistant, F-"
/clone_idb="Mouse 10kb plasmid cDNA library"
/notes="Vector: pMD42nt; Purified genomic DNA from M."

```

muscularis C57BL/6J (female) was obtained from M. Jackson Laboratory (Bar Harbor, ME). DNA Resource (<http://www.jax.org/resources/documents/dnarses/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to

10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pMD2 (g14732114|gbl|AF1902.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adapted vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance."

ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 669;
Best Local Similarity 54.3%; Pred. No. 97;
Matches 69; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 121 CCGACTACAGCTGATGAGTATGATGAGTCCGCTACGCTGCTGCTGCA 180
Db 235 CTGAGGAGATGCTTAAAGGTTTCAGAGCAATGTAAGTCCATGCTTCTGCTT 176
Qy 181 AGCGATGCCGTAATATGATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCT 240
Db 175 CCGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 116
Qy 241 GCAGGCC 247
Db 115 CATGGAC 109

RESULT 8
LOCUS BZ569281
DEFINITION BZ569281 978 bp DNA linear GSS 17-DEC-2002
ACCESION BZ569281
VERSION BZ569281
KEYWORDS BZ569281.1 GI:27203648
SOURCE GSS.
ORGANISM *Pseudomonas aeruginosa*
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.

REFERENCE 1 (bases 1 to 978)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M., Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of *Pseudomonas aeruginosa* library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.

FEATURES
source 1..978
Location/Qualifiers
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="2-164"
/db_xref="taxon:287"
/clone="pacs2-164 8328"
/clone_lib="pacs2-164"
/note="Clinical isolate 2-164 whole genomic shotgun library."

ORIGIN

Query Match 7.5%; Score 34.2; DB 28; Length 978;
Best Local Similarity 53.3%; Pred. No. 1.1e+02;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 213 CCGTGTGTTACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 272
Db 519 CGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 578

Qy 273 ACTGACTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 332
Db 579 GATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 638
Qy 333 CGATATTACTGCTCGG 347
Db 639 TGTAGTACCGGCGAG 653

RESULT 9
LOCUS BG671547
DEFINITION BG671547 1036 bp mRNA linear EST 30-APR-2001
DNRBUB05 Rat DRG Library Rattus norvegicus cDNA clone DNRBUB05 5',
mRNA sequence.
VERSION BG671547
KEYWORDS BG671547.1 GI:13893646
EST.

SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1036)
AUTHORS Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L., Huang,W.J., Pu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang,X., Chen,Z., Han,Z.G. and Zhang,X.
TITLE Identification of gene expression profile of dorsal root ganglion in the rat peripheral axotomy model of neuropathic pain
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
MEDLINE 22056133
PubMed 12060780
COMMENT Contact: Zhang Xu
Laboratory of Sensory System
Institute of Neuroscience
320 Yue Yang Road, Shanghai 200031, P.R.China
Tel: 86-21-64748700-121
Fax: 86-21-64713446
Email: xu.zhang@ion.ac.cn
This clone is also available at Chinese National Human Genome Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong New Area, P.R.China. Please contact with Zhang Xu (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
PCR Primers
FORWARD: T3
BACKWARD: T7
Seq primer: T3
POLYA=No.

FEATURES
source 1..1036
Location/Qualifiers
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="DNRBUB05"
/sex="male"
/tissue_type="dorsal root ganglion"
/dev_stage="adult"
/clone_lib="Rat DRG Library"

ORIGIN

Query Match 7.5%; Score 34.2; DB 12; Length 1036;
Best Local Similarity 49.2%; Pred. No. 1.2e+02;
Matches 90; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

Qy 11 TAAAGTGGAGCATGCTGAGCAATGCTAGTCTGAGAGTCTGCTGCTGCTGCTGCT 70
Db 168 TGAAGATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 227
Qy 71 CACATGG 130
Db 228 CTGTAGAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 287

mus musculus (house mouse)

mus musculus (house mouse)

7.58; Score 34; DB 10; Length 234.

Beet Local Similarity 51.3%; Pred. No. 70;
Matches 79; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

Qy 35 TCGTAGTTTGGGAGTGTCTGCTGGCGCTTCACATAGGGCGCGGTAATC 94
Db 200 TAGTTTGGTAGAGTGTGCTTTATGAGGGGTTCTTTCACAGCGCCGCGCTTCC 141

Qy 95 ATAACGGGGGGGCAATAGTTCCGGCCGGACTCAGCTGAGCATTTATAGTAGTGGTT 154
Db 140 CGAACCGCGGCTCTTTATAGCTTTGTTCAAAATGTGAGGCGCATTTACACACCTTT 81

Qy 155 CCGCTACGCTGCGCTGCTGCTGCAAGCATGC 188
Db 80 CCGTTACCAAGCCCGTGGCGCCCATTCGCGCGC 47

RESULT 12
CA711477 500 bp mRNA linear EST 26-NOV-2002
LOCUS wdk2c.pk014.03 wdk2c Triticum aestivum cDNA clone wdk2c.pk014.03 5'
DEFINITION end, mRNA sequence.
ACCESSION CA711477.1 GI:25433270
VERSION CA711477.1 GI:25433270
KEYWORDS EST.
SOURCE Triticum aestivum (bread wheat)
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Triticum.
REFERENCE 1 (bases 1 to 500)
Tingley,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE Dupont Wheat cDNA Sequence
JOURNAL Unpublished (2002)
COMMENT Contact: Scott V. Tingley
Crop Genetics
E. I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingley@USA.dupont.com
Seq primer: M13.

FEATURES
source 1..500
Location/Qualifiers
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdk2c.pk014.03"
/tissue_type="kernel"
/clone_lib="wdk2c"
/note="Vector: pBluescript SK+, Site_1: EcoRI; Site_2:
XhoI; Wheat (Triticum aestivum L.) developing kernel, 7
days after anthesis."

ORIGIN
Query Match 7.5%; Score 34; DB 14; Length 500;
Best Local Similarity 54.5%; Pred. No. 97;
Matches 61; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 15 AGTGCAGCATTTGCGAGCATGTTGTCAGTGTCTGCTGCGCTGCTTCACA 74
Db 274 AGTGNATCAGATCGCAGCCGCGCTCCCGCGCTGTCGAACCGCGGAGCTCCCN 215

Qy 75 ATGGGGCGCGCGGTATATCATACGCGCGCGCATATGTTCCGCGCGGAC 126
Db 214 TGTGGAGCGGTGGCATGACATGCGCGCGCGCGCGGTGGCGCCAC 163

RESULT 13
LOCUS AZ180696 500 bp DNA linear GSS 30-AUG-2000
DEFINITION SP_0181_A1_c06 SP6E Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library Strongylocentrotus purpuratus

genomic clone Plate=181 Col=11 Row=E; genomic survey sequence.
AZ180696
AZ180696.1 GI:8353071
GSS.
Strongylocentrotus purpuratus
Strongylocentrotus purpuratus
Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
Echinoidae; Euechinozoa; Echinacea; Echinoida;
Strongylocentrotidae; Strongylocentrotus.
REFERENCE 1 (bases 1 to 500)
Cameron,R.A., Mahatras,G., Rast,J.P., Martinez,P., Biondi,T.R.,
Swartzell,S., Wallace,J.C., Pousetka,A.J., Livingston,B.T.,
Wray,G.A., Betensohn,C.A., Lebrach,H., Britten,R.J., Davidson,E.H.
and Hood,L.
A sea urchin genome project: Sequence scan, virtual map, and
additional resources
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
MEDLINE 20402566
PUBMED 10920195
COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
Division of Biology 156-29
California Institute of Technology
Pasadena California 91125, USA
Tel: (626) 395-8421
Fax: (626) 793-3047
Email: acameron@caltech.edu
Plate: 181 Row: E Column: 11
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 500.
Location/Qualifiers
1..500
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="Plate=181 Col=11 Row=E"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC6.6; BAC clones in E-Coli
DH10B"

FEATURES
source 1..500
Location/Qualifiers
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="Plate=181 Col=11 Row=E"
/clone_lib="Strongylocentrotus purpuratus, purple sea
urchin, sperm genomic BAC library"
/note="Organ: sperm; Vector: BAC6.6; BAC clones in E-Coli
DH10B"

ORIGIN
Query Match 7.5%; Score 34; DB 28; Length 500;
Best Local Similarity 50.3%; Pred. No. 97;
Matches 82; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

Qy 133 TTGAGCATTTATCAGTACGCTTCGCTGCAAGGATGCCGT 192
Db 8 TTATCCATATATATCCATGATTAACGATGATGATGATGATGATGATGA 67

Qy 193 AAATATGATCAGCTGTTACCCGCTGTTTACCATGAATGCGACATGCGCAGGCT 252
Db 68 GACCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 127

Qy 253 GCGATATAGTACTATTGAACTGACTCAGAAAGTTCGAA 295
Db 128 GATATGATATATATATGAAATTAATTAATGATATACGAA 170

RESULT 14
LOCUS AA746477 365 bp mRNA linear EST 27-JAN-1998
DEFINITION NW62802.81 NCI CGAP GCB1 Homo sapiens cDNA clone IMAGE:1251194 3'
similar to TR:000555 000555 NEURONAL CALCIUM CHANNEL ALPHA 1A
SUBUNIT ISOFORM A-1. ; contains element TARI repetitive element ;
mRNA sequence.
ACCESSION AA746477
VERSION AA746477.1 GI:2786463
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 365)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bernaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 found through the I.M.A.G.E. Consortium/LINL at:
 www-bio.llnl.gov/bdip/image/image.html
 Insert Length: 661 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from AmerSham
 High quality sequence stop: 292.
 Location/Qualifiers

FEATURES
 source 1..365
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1251194"
 /issue_type="germline center B cell"
 /lab_host="DH10B"
 /clone_lib="NCI CGAP GC81"
 /note="Vector: p773D-Pac (pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD10+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CIBR). cDNA synthesis was
 primed with a Not I - oligo(dT) primer:
 5'-TGTTACCAATCTGAGAGGAGGCGCGCTCTTTT-3'
 1) Double-stranded cDNA was ligated to Eco RI adaptors
 (pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified p773 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bernaldo."

ORIGIN

Query Match 7.4%; Score 33.8; DB 9; Length 365;
 Best Local Similarity 56.9%; Pred. No. 97;
 Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 Oy 15 AGTGCACATTCGACGACATCGTAGTTCTGCGAGTCTCGCGCGCTGTCACCA 74
 Db 119 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCGGCGCGCGCGCGCGCGCG 60
 Oy 75 ATGGGCGCGCGCGCGGTATCTTAACGCGCGCGCGCATAGTTCGCGCCG 123
 Db 59 AGCGCCCTCGAGGTACCGCAGCGCCGCGCGCGCATAGTTCGCGCCG 11

RESULT 15
 CD374421/c 558 bp mRNA linear EST 19-AUG-2003
 LOCUS TWBm1C2H3T3 Wuchereria bancrofti microfilaria cDNA (SAM953JL-WDMF)
 DEFINITION Wuchereria bancrofti cDNA clone TWBm1C2H3 5', mRNA sequence.
 ACCESSION CD374421
 VERSION CD374421.1 GI:31229939
 KEYWORDS EST.
 SOURCE Wuchereria bancrofti
 ORGANISM Wuchereria bancrofti
 Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidae;
 Onchocercidae; Wuchereria.

REFERENCE 1 (bases 1 to 558)
 AUTHORS Ndi, J.S., Ribeiro, J.M. and Nutman, T.B.
 TITLE Toward the characterization of the Wuchereria bancrofti
 microfilarial transcriptome with comparisons to those of Brugia
 malayi and Onchocerca volvulus

JOURNAL

COMMENT Unpublished (2003)
 Contact: Thomas B. Nutman
 Laboratory of Parasitic Diseases
 NIH
 Building 4 Room 126, Bethesda, MD 20892-0425, USA
 Email: tnutman@niaid.nih.gov
 Seq primer: pBluescript T3.

FEATURES
 source 1..558
 Location/Qualifiers

/organism="Wuchereria bancrofti"
 /mol_type="mRNA"
 /db_xref="taxon:6293"
 /clone="TWBm1C2H3"
 /dev_stage="microfilaria"
 /lab_host="X11-Blue MRF"
 /clone_lib="Wuchereria bancrofti microfilaria cDNA
 (SAM953JL-WDMF)"
 /note="Vector: Lambda Uni-ZAP XR; Site_1: Eco RI; Site_2:
 Xho I; Lymphatic filarial nematode parasite of humans.
 mRNA was prepared from approximately 85,000 microfilariae
 isolated from the blood of an infected individual from
 Guyana, South America and converted to double-stranded
 cDNA using reverse transcriptase and oligo(dT) followed by
 Kase H and DNA pol I. The library has 2.9 x 10⁵
 independent recombinants and the average insert size is
 ~1kb. The library was constructed by Sandra J. Laney. The
 library is available from Dr.S.A.Williams, email:
 genome@smith.edu."

ORIGIN

Query Match 7.4%; Score 33.8; DB 14; Length 558;
 Best Local Similarity 60.7%; Pred. No. 1.2e+02;
 Matches 65; Conservative 0; Mismatches 40; Indels 2; Gaps 1;
 Oy 76 TGCGCGCGCGCGGTATCATTAAGCGCGCGCAATAGTCCGCCGCACTCAAGTTG 135
 Db 161 TGGCGCGCGCGGTATCAACCGGTGCGGTATATATTAAGCGCGCAATTCATGCT 102
 Oy 136 AGCATTATC--AGTACGTTCCGCTAAGCTGCGCTGCTGCA 180
 Db 101 GACATTATCAAGANCGTGGCGGANNCTGTGCTGCTGCTGNA 55

Search completed: March 16, 2004, 04:29:16
 Job time : 2234.91 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using BW model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds

(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456
Sequence: 1 atgaactcttaaaagtcgsc.....ccacgcgtacacagatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA Main:*

```

1: /cgn2_6/ptodata/2/pna/US06_PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*

```

```

44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US105_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	US-08-233-642A-56	Sequence 56, Appl
2	456	100.0	456	US-09-543-407-1	Sequence 1, Appl
3	412.8	90.5	456	US-09-543-407-19	Sequence 19, Appl
4	406.4	89.1	456	US-09-543-407-17	Sequence 17, Appl
5	404.8	88.8	456	US-09-543-407-11	Sequence 11, Appl
6	404.8	88.8	456	US-09-543-407-13	Sequence 13, Appl
7	404.8	88.8	456	US-09-543-407-23	Sequence 23, Appl
8	404.8	88.8	456	US-09-543-407-25	Sequence 25, Appl
9	403.2	88.4	456	US-09-543-407-15	Sequence 15, Appl
10	401.6	88.1	456	US-09-543-407-21	Sequence 21, Appl
11	401.6	88.1	456	US-09-543-407-27	Sequence 27, Appl
12	398.4	87.4	456	US-09-543-407-29	Sequence 29, Appl
13	331.8	72.8	361	US-08-233-642A-54	Sequence 54, Appl
14	254.4	55.8	456	US-09-543-407-3	Sequence 3, Appl
15	254.4	55.8	456	US-09-543-407-172	Sequence 172, Appl
16	254.4	55.8	477	US-09-252-691-172	Sequence 172, Appl
17	254.4	55.8	477	US-10-146-492B-172	Sequence 172, Appl
18	254.4	55.8	675	US-08-978-878-3	Sequence 3, Appl
19	254.4	55.8	675	US-09-741-873B-3	Sequence 3, Appl
20	252	55.3	453	US-60-352-946-1	Sequence 1, Appl
21	252	55.3	453	US-60-444-371-1	Sequence 1, Appl
22	210.4	46.1	336	US-08-978-878-1	Sequence 1, Appl
23	185.6	40.7	360	US-09-741-873B-1	Sequence 1, Appl
24	64.8	14.2	522	US-09-252-691-170	Sequence 170, Appl
25	64.8	14.2	522	US-10-146-492B-170	Sequence 170, Appl
26	64.8	14.2	522	US-10-146-492B-170	Sequence 170, Appl
27	51.2	11.2	100	US-10-417-886-171	Sequence 171, Appl
28	44	9.6	200	US-09-887-272A-5263	Sequence 5263, Appl
29	42.4	9.3	100	US-10-146-492B-79	Sequence 79, Appl
30	40.8	8.9	100	US-10-146-492B-79	Sequence 79, Appl
31	37.4	8.2	1304	US-10-437-963-19773	Sequence 19773, Appl
32	35.8	7.9	561	US-09-252-691-171	Sequence 171, Appl
33	35.8	7.9	561	US-09-252-691-171	Sequence 171, Appl
34	35.8	7.9	561	US-10-417-886-171	Sequence 171, Appl
35	35.4	7.8	3411	PCT-US01-08631-24330	Sequence 24330, A
36	35.4	7.8	3411	PCT-US01-08631-24330	Sequence 24330, A
37	34.6	7.6	438	US-09-739-449-3383	Sequence 3383, Appl
38	34.6	7.6	438	US-09-803-110-3383	Sequence 3383, Appl
39	34.6	7.6	456	US-09-543-407-4	Sequence 4, Appl
40	34.6	7.6	29058	US-09-514-000-342	Sequence 342, Appl
41	34.6	7.6	187892	US-09-739-449-208	Sequence 208, Appl
42	34.6	7.6	187892	US-09-803-110-208	Sequence 208, Appl
43	34.2	7.5	385	US-09-333-535A-288	Sequence 288, Appl
44	34.2	7.5	385	US-09-615-606A-909	Sequence 909, Appl
45	34.2	7.5	396	US-60-089-524-4522	Sequence 4522, Appl

ALIGNMENTS

RESULT 1
US-08-233-642A-56
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESSES:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043, 403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 100.0%; Score 456; DB 6; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.9e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAACCTTTAAAGTGCGAGCATTCGACCAATCGTAGTTTCGAGTCTGCTGCT 60
1 ATGAACCTTTAAAGTGCGAGCATTCGACCAATCGTAGTTTCGAGTCTGCTGCT 60
61 GGGCGTGTTCACCATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
61 GGGCGTGTTCACCATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 120
121 CGGACTCAAGTGTGAGCATTTATCAGTAGCGGTTACGCTGAGCGTGTGCTGCGCA 180
121 CGGACTCAAGTGTGAGCATTTATCAGTAGCGGTTACGCTGAGCGTGTGCTGCGCA 180
121 AGGATGCGGCTTAATCTGAACGACCATTAACGAGCGGTTATGTTAAGCGCGCAT 240
121 AGGATGCGGCTTAATCTGAACGACCATTAACGAGCGGTTATGTTAAGCGCGCAT 240
121 AGGATGCGGCTTAATCTGAACGACCATTAACGAGCGGTTATGTTAAGCGCGCAT 240
241 GTAGGCGGAGGTGGGATTAATGACTATGAACTGAGTGAAGTGTTCAGAAATTAAT 300
241 GTAGGCGGAGGTGGGATTAATGACTATGAACTGAGTGAAGTGTTCAGAAATTAAT 300
301 GCCACATTCGACGATGGAAGCGTAAAACTCGGATTAATGCTGCGCAATACGCGGT 360
301 GCCACATTCGACGATGGAAGCGTAAAACTCGGATTAATGCTGCGCAATACGCGGT 360
361 AATAAGCCCGCGTGTGTTATCAGACCGCATTCGATTCAGAGGTTATGTTGCTGAGTT 420
361 AATAAGCCCGCGTGTGTTATCAGACCGCATTCGATTCAGAGGTTATGTTGCTGAGTT 420
421 GGTGTTGGCAACGCGGCTTAACGAGTTTAA 456
421 GGTGTTGGCAACGCGGCTTAACGAGTTTAA 456

RESULT 2
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.

;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543,407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO: 1
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: *Salmonella enteritidis*
US-09-543-407-1

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2,9e-115;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGCGGTAAATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACAATGGGGCGGCGCGGTAAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAAGCGCGCAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAAGCGCGCAT 240
Qy 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAATGACTGACTGAGATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATAGTACTATTTGAATGACTGACTGAGATGTTTCAAAATAT 300
Qy 301 GCCACATCGACGAGTGAACGCTTAAATCTCCGATATTACTGTGCGGCATACGCGCGT 360
Db 301 GCCACATCGACGAGTGAACGCTTAAATCTCCGATATTACTGTGCGGCATACGCGCGT 360
Qy 361 AATAAGCGCGGCTGTAAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 420
Db 361 AATAAGCGCGGCTGTAAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 420
Qy 421 GGTTTGGCAACAACGCCACGCGCTAACCGATATTAA 456
Db 421 GGTTTGGCAACAACGCCACGCGCTAACCGATATTAA 456

RESULT 3
US-09-543-407-19
; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.

US-09-543-407-19

Query Match 90.5%; Score 412.8; DB 23; Length 456;
Best Local Similarity 94.1%; Pred. No. 2,2e-121;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGCGGTAAATCAATACGCGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCACAATGGGGCGGCGCGGTAAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATAGTACGCTTCCTGCAACGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAAGCGCGCAT 240
Db 181 AGCGATGCCCGTAAATCTGAAACGACATTAACCGAGCGGTTATGTAAAGCGCGCAT 240
Qy 241 GTAGGCCAGGGTGGCGATTAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 300
Qy 301 GCCACATCGACGAGTGAACGCTTAAATCTCCGATATTACTGTGCGGCATACGCGCGT 360
Db 301 GCCACATCGACGAGTGAACGCTTAAATCTCCGATATTACTGTGCGGCATACGCGCGT 360
Qy 361 AATAAGCGCGGCTGTAAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 420
Db 361 AATAAGCGCGGCTGTAAATAGTACTGATTTGAACTGACTGAGATGTTTCAAAATAT 420
Qy 421 GGTTTGGCAACAACGCCACGCGCTAACCGATATTAA 456
Db 421 GGTTTGGCAACAACGCCACGCGCTAACCGATATTAA 456

RESULT 4
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-17

Query Match 89.1%; Score 406.4; DB 23; Length 456;
Best Local Similarity 93.2%; Pred. No. 2,5e-119;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Db 1 ATGAAACTTTTAAAGTGGCAGCATTCGAGCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
Qy 61 GCGCTGTTCCACAATGGGGCGGCGCGGTAAATCAATACGCGCGGCAATAGTTCCGGC 120

```

Db      61  GGCCTCGTTCCAAAGTGGGCGCGGCTAATACAGCGCGGCAATAGTTCCGGC 120
Qy      121  CCGGACTCAAGTGTGACATTTATCAGTACGGTTCCGCTAACGCTGCTGTCTGCAA 180
Db      121  CCGGACTATGATCAGCTGGTTACCGGTGTGTACCCATGAAATGAGACATGACCTGCAA 180
Qy      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Db      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Qy      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGACTGAAATGTTTCAAGAAATAT 300
Db      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Qy      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Db      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Qy      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Db      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Qy      421  GGTTCGCAACCAACCGCGCTAATCCAGTATTA 456
Db      421  GGTTCGCAACCAACCGCGCTAATCCAGTATTA 456

```

RESULT 5

```

US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-11

```

```

Query Match      88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

```

Qy      1  ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTTGGCAGTCTCTGCT 60
Db      1  ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTTGGCAGTCTCTGCT 60
Qy      61  GGCCTCGTTCCAAAGTGGGCGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Db      61  GGCCTCGTTCCAAAGTGGGCGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy      121  CCGGACTCAAGTGTGACATTTATCAGTACGGTTCCGCTAACGCTGCTGTCTGCAA 180
Db      121  CCGGACTCAAGTGTGACATTTATCAGTACGGTTCCGCTAACGCTGCTGTCTGCAA 180
Qy      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Db      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Qy      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Db      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Qy      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Db      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Qy      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Db      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420

```

```

Qy      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Db      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Qy      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Db      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Qy      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Db      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Qy      421  GGTTCGCAACCAACCGCGCTAATCCAGTATTA 456
Db      421  GGTTCGCAACCAACCGCGCTAATCCAGTATTA 456

```

RESULT 6

```

US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; FILE REFERENCE: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-13

```

```

Query Match      88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

```

Qy      1  ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTTGGCAGTCTCTGCT 60
Db      1  ATGAACCTTTTAAAGTGGAGCATTCGACGATCGTATGTTCTTGGCAGTCTCTGCT 60
Qy      61  GGCCTCGTTCCAAAGTGGGCGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Db      61  GGCCTCGTTCCAAAGTGGGCGCGGCTAATCAATACGCGCGGCAATAGTTCCGGC 120
Qy      121  CCGGACTCAAGTGTGACATTTATCAGTACGGTTCCGCTAACGCTGCTGTCTGCAA 180
Db      121  CCGGACTCAAGTGTGACATTTATCAGTACGGTTCCGCTAACGCTGCTGTCTGCAA 180
Qy      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Db      181  AGCGATGCCCGTAAATCTGAAACGACATTCACAGACGGGTATGTTAAGCGCGCAT 240
Qy      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Db      241  GTAGGCGAGGGTGGCGATTAATAGTACTATTAAGTACTGAAATGTTTCAAGAAATAT 300
Qy      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Db      301  GGCACCATGACACGATGGAACGCTAAACCTCCGATTAATCTGTCCGCAATAGCGCGGT 360
Qy      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420
Db      361  AATAACCGCGCGCTGTTAATCAGACCGCATCTGATTCAGGCTAATGCTGCTCAGGTT 420

```


QY 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456
DB 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456

RESULT 7

US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 88.8%; Score 404.8; DB 23; Length 456;

Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
QY 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGTAACGTCGCTGCTGCAA 180
DB 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGTAACGTCGCTGCTGCAA 180
QY 181 AGCGATCCCGTAAATCTGAACGACATTAACGAGCGGTTATGTAAACGGCGCGAT 240
DB 181 AGCGATCCCGTAAATCTGAACGACATTAACGAGCGGTTATGTAAACGGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATAGTACTTGAACCTGACTCAGAAATGTTTCAAAATAT 300
DB 241 GTAGCCAGGGTGGGATTAATAGTACTTGAACCTGACTCAGAAATGTTTCAAAATAT 300
QY 301 GCCACATCGACGAGTGAACGCTAAAACTCCGATTAATCTGTCGGCATAAGCGCGT 360
DB 301 GCCACATCGACGAGTGAACGCTAAAACTCCGATTAATCTGTCGGCATAAGCGCGT 360
QY 361 AATAAGCCGGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAATGGTCCGTCAGGTT 420
DB 361 AATAAGCCGGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAATGGTCCGTCAGGTT 420
QY 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456
DB 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456

RESULT 8

US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-25

Query Match 88.8%; Score 404.8; DB 23; Length 456;

Best Local Similarity 93.0%; Pred. No. 8.3e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCAATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCAATTCGACCAATCGAGTTCTGGAGCTCTGGCT 60
QY 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGTAACGTCGCTGCTGCAA 180
DB 121 CCGACTCAACGTTGAGCATTTATCAGTACGTTCCGTAACGTCGCTGCTGCAA 180
QY 181 AGCGATCCCGTAAATCTGAACGACATTAACGAGCGGTTATGTAAACGGCGCGAT 240
DB 181 AGCGATCCCGTAAATCTGAACGACATTAACGAGCGGTTATGTAAACGGCGCGAT 240
QY 241 GTAGCCAGGGTGGGATTAATAGTACTTGAACCTGACTCAGAAATGTTTCAAAATAT 300
DB 241 GTAGCCAGGGTGGGATTAATAGTACTTGAACCTGACTCAGAAATGTTTCAAAATAT 300
QY 301 GCCACATCGACGAGTGAACGCTAAAACTCCGATTAATCTGTCGGCATAAGCGCGT 360
DB 301 GCCACATCGACGAGTGAACGCTAAAACTCCGATTAATCTGTCGGCATAAGCGCGT 360
QY 361 AATAAGCCGGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAATGGTCCGTCAGGTT 420
DB 361 AATAAGCCGGCGGTGTTAATCAGACCGCATCTGATTCAGAGTAATGGTCCGTCAGGTT 420
QY 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456
DB 421 GGTITGGCAACAACGCCAGCTTAACAGATTAA 456

RESULT 9

US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of Leishmania major.
US-09-543-407-15

Query Match 88.1%; Score 403.2; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 2,7e-118;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180
DB 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180
QY 181 AGCGATGCCGTAATCTGAAAGCAATTAACGAGGCGGTTATGTAAGCGGCGGAT 240
DB 181 AGCGATGCCGTAATCTGAAAGCAATTAACGAGGCGGTTATGTAAGCGGCGGAT 240
QY 241 GTAGGCGAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 GTAGGCGAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GGCACCATGACGAGTGAACGCTAAACCTCCGATTTATGTCGCGCAATACGGCGT 360
DB 301 GGCACCATGACGAGTGAACGCTAAACCTCCGATTTATGTCGCGCAATACGGCGT 360
QY 361 AATAAGCGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 AATAAGCGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GGTITGGCAACGAGCGGCTAATGATGATGATGATGATGATGATGATGATG 456
DB 421 GGTITGGCAACGAGCGGCTAATGATGATGATGATGATGATGATGATGATG 456

RESULT 10
US-09-543-407-21

Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of Leishmania major.
US-09-543-407-21

Query Match 88.1%; Score 401.6; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 8.8e-118;

Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180
DB 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180
QY 181 AGCGATGCCGTAATCTGAAAGCAATTAACGAGGCGGTTATGTAAGCGGCGGAT 240
DB 181 AGCGATGCCGTAATCTGAAAGCAATTAACGAGGCGGTTATGTAAGCGGCGGAT 240
QY 241 GTAGGCGAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGATG 300
DB 241 GTAGGCGAGGTCGGATTAATGATGATGATGATGATGATGATGATGATGATG 300
QY 301 GGCACCATGACGAGTGAACGCTAAACCTCCGATTTATGTCGCGCAATACGGCGT 360
DB 301 GGCACCATGACGAGTGAACGCTAAACCTCCGATTTATGTCGCGCAATACGGCGT 360
QY 361 AATAAGCGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 AATAAGCGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GGTITGGCAACGAGCGGCTAATGATGATGATGATGATGATGATGATGATG 456
DB 421 GGTITGGCAACGAGCGGCTAATGATGATGATGATGATGATGATGATGATG 456

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from Gpe3 of Leishmania major.
US-09-543-407-27

Query Match 88.1%; Score 401.6; DB 23; Length 456;
Best Local Similarity 92.5%; Pred. No. 8.8e-118;
Matches 422; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGAGCATTCGAGCAATGATGTTCTGCGCTGCTGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180
DB 121 CCGGACTCAACGTTGAGCATTTATGATGATGATGATGATGATGATGATGATG 180

Db 121 CGGAGCTCAGGTTGAGCATTTATCATGTAAGGTTCCGTAACGCTGCGCTTCTCTGCA 180
Qy 181 AGCGATGCGCGTAAATCTGAAAACGACATTCACAGACGGTATGTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAAACGACATTCACAGACGGTATGTATGTAACGGCGCGAT 240
Qy 241 GTAGGCGCAGGATGCGGATTAATGTAAGTGAAGTCAAGTGAAGTTCAGAAATAT 300
Db 241 TATATCTAGCTGTTTACCGCTGTGTGTTACCGTAAATGGCACATTCAGAAATAT 300
Qy 301 GCCACATCGACAGTGAAGCGTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCCACATCGACAGTGAAGCGTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Db 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Qy 421 GGTTTGGCAACAACGCCAGCGCTAACAGTATTA 456
Db 421 GGTTTGGCAACAACGCCAGCGCTAACAGTATTA 456

RESULT 12
US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: Parseq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-29

Query Match 87.4%; Score 398.4; DB 23; Length 456;
Best Local Similarity 92.1%; Pred. No. 9.4e-117;
Matches 420; Conservative 0; Mismatches 36; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACAGCATTCGAGTTCGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACAGCATTCGAGTTCGAGTCTGCTGCT 60
Qy 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Db 61 GGCCTGTTCCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGCTTGAAGATTTATCAGTACGCTTCGCTACGCTGCTCTCTGCA 180
Db 121 CCGGACTCAAGCTTGAAGATTTATCAGTACGCTTCGCTACGCTGCTCTCTGCA 180
Qy 181 AGCGATGCGCGTAAATCTGAAAACGACATTCACAGACGGTATGTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAAACGACATTCACAGACGGTATGTATGTAACGGCGCGAT 240
Qy 241 GTAGGCGCAGGATGCGGATTAATGTAAGTGAAGTCAAGTGAAGTTCAGAAATAT 300
Db 241 GTAGGCGCAGGATGCGGATTAATGTAAGTGAAGTCAAGTGAAGTTCAGAAATAT 300

Qy 301 GCCACATCGACAGTGAAGCGTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCCACATCGACAGTGAAGCGTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Db 361 AATAACGCGCGCTGTTAATCAGACCGCATCTGATTCAGAGTAATGTGCTCAGGTT 420
Qy 421 GGTTTGGCAACAACGCCAGCGCTAACAGTATTA 456
Db 421 GGTTTGGCAACAACGCCAGCGCTAACAGTATTA 456

RESULT 13
US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1...357
US-08-233-642A-54

Query Match 72.8%; Score 331.8; DB 6; Length 361;
Best Local Similarity 99.4%; Pred. No. 2.2e-95;
Matches 333; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 64 GTGCTTCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGCGCG 123
Db 1 GTGCTTCACAATGGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGGCGCG 60
Qy 124 GACTCAACGTTGAGCATTTTATCAGTACGCTTCGCTAACGCTGCGCTCTGCAAGC 183
Db 61 GACTCAACGTTGAGCATTTTATCAGTACGCTTCGCTAACGCTGCGCTCTGCAAGC 120

Search completed: March 18, 2004, 02:46:32
Job time : 2926.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69 6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-1

Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgctaccagtactataa 456

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA New:
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	254.4	55.8	675	US-09-741-873C-3	Sequence 3, Appli
2	210.4	46.1	396	US-09-741-873C-1	Sequence 1, Appli
3	33.4	7.3	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-19A-1	Sequence 1, Appli
5	31.4	6.9	1588	US-10-767-701-14522	Sequence 14522, A
6	30.8	6.8	1575	US-10-788-792-27	Sequence 27, Appli
7	29.6	6.5	24900	US-10-767-471-10736	Sequence 10736, A
8	29.4	6.4	300	US-10-767-701-28437	Sequence 28437, A
9	29.4	6.4	1355	US-10-045-674A-593	Sequence 593, App
10	29.4	6.4	1400	US-60-545-213-2134	Sequence 2134, Ap
11	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, Ap
12	29.4	6.4	1400	US-60-545-213-6406	Sequence 6406, Ap
13	29.4	6.4	1400	US-60-545-213-6407	Sequence 6407, Ap
14	29.2	6.4	249	US-10-771-241-117	Sequence 117, Appli
15	29.2	6.4	757	US-10-771-241-29	Sequence 29, Appli
16	28.8	6.3	83009	US-10-417-375A-143	Sequence 143, App
17	28.6	6.3	1186	PCT-US04-05654-1030	Sequence 1030, Ap
18	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, Ap
19	28.2	6.2	201	US-60-548-091-22688	Sequence 22688, A
20	28.2	6.2	596	US-10-767-701-5378	Sequence 5378, Ap
21	28.2	6.2	610	US-10-767-701-8781	Sequence 8781, Ap
22	28.2	6.2	1295	US-10-100-683-5480	Sequence 5480, Ap
23	28.2	6.2	1746	PCT-US04-05654-2555	Sequence 2555, Ap
24	28.2	6.2	3416	US-10-100-683-3397	Sequence 3397, Ap
25	28.2	6.2	8206	US-10-021-698A-3529	Sequence 3529, Ap
26	28.2	6.1	237	US-10-767-701-31417	Sequence 31417, A

c	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
c	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
c	29	27.8	6.1	1749	6	US-10-767-701-12516	Sequence 12516, A
c	30	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
c	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
c	32	27.6	6.1	600	7	US-60-545-213-6290	Sequence 2020, Ap
c	33	27.6	6.1	1052	7	US-60-545-213-6292	Sequence 6292, Ap
c	34	27.6	6.1	1055	6	US-10-767-701-12607	Sequence 12607, A
c	35	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
c	36	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
c	37	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
c	38	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
c	39	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
c	40	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, Ap
c	41	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
c	42	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, Ap
c	43	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
c	44	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
c	45	27.2	6.0	2880	6	US-10-100-683-5057	Sequence 5057, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Staffan
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 06/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 06/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 06/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 55.8%; Score 254.4; DB 5; Length 675;
Best Local Similarity 72.4%; Pred. No. 2.5e-77;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
Qy 1 ATGAACCTTTAAAGTGCAGCAATTCGATTTCTGCGAGTCTGAGCT 60
Db 83 ATGAACCTTTAAAGTGCAGCAATTCGATTTCTGCGAGTCTGAGCT 142
Qy 61 GCGCTGCTTCAATGCGGCGGCGGTAATCATACGCGCGGCAATGTTCCGGC 120
Db 143 GGTGTTTCTCTGTCGCGGCGGCGGTAATCATACGCGCGGCAATGTTCCGGC 202
Qy 121 CCGACTCAACGTTGACATTTATCAGTACGCTTCCGCTAACGCTGCGCTTGCCTGCA 180

```
Db 203 CCAATTCGAGCTGACATTTACCATGACGGTGGGTAACCTGCACTTGTCTGCAA 262
Qy 181 AGCGATGCCCGCTTAATCTGAAACGACATTAACCGAGCGGTTATGGTAAACGGCGCGAT 240
Db 263 ACTGATGCGCCCTTAACCTGACTGACTGATCTATACCCAGATGGCGCGGTTATGGTCAAGAT 322
Qy 241 GTAGCCGCGGCTGGGATATAGTAACTGAACTGAACTGAACTGAACTGAACTGAACTGAA 300
Db 323 GTTGTGACGGGCTGAGATGACAGCTCAATGATGACCAAGTGGCTTGGTAAACAGC 382
Qy 301 GCCACATCGACCGCTGAGTGAACGCTTAAACTCCGATATTAAGTCCGCAATACGGCGGT 360
Db 383 GCTACTCTTGAATGAGTGAACGCGCAAAATTTGAAATGAGGTTTAAACAGTTCCGTTGAT 442
Qy 361 AATAACCGCGGCTGGTAAATGACAGCGCATGATTCAGCGTAAATGGTGGTCAAGTT 420
Db 443 GCGAAGCGTGTGAGTGAAGTTGACCAAGCTGACATCTCACTCCCTCCGTCAGTCACTGAGTT 502
Qy 421 GGTTTGGCAACAACGCGCAGCGGCTTAACCACTATTAA 456
Db 503 GGGTTTGGTAAACAACGCGCAGCGCTCACTCACTAA 538
```

RESULT 2

```
US-09-741-873C-1
Sequence 1, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normak, Staffan
FILE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873C-1
```

```
Query Match 46.1%; Score 210.4; DB 5; Length 396;
Best Local Similarity 70.7%; Pred. No. 2.2e-62;
Matches 280; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
```

```
Qy 61 GGCCTGCTTCCACAAATGGCGCGCGGATATCATTAACGGCGCGCAATATGTCGCGC 120
Db 1 GGTGTTCTTCTCAATGACGGCGCGGATTAACAGCGTGGCGGTAATATAGCGGC 60
Qy 121 CCGGACTCAACGTTGAGCATTTATGATCGGTTCCGTAAGCGTGGCTGCTTGCNA 180
Db 61 CCAAAATTCGAGCTGAACATTTACGATACGGTGGGATCTTGCACCTGCTTGCNA 120
Qy 181 AGGATGCGCGGATATCTGAACGACATTAACCAAGCGGTTATGATGATGCGCGCGAT 240
Db 121 ACTGATGCGCGGATCTGATCTGATTAACCAAGCATGGCGGATATGATGCGAGAT 180
Qy 241 GTAGCGCAGGCTGGGATATAGTAACTGAACTGAACTGAACTGAACTGAACTGAA 300
```

```
Db 181 GTTGTGAGGCTGATGATGACAGCTCAATGATCTGACCAAGGTGGCTTGGTAAACAGC 240
Qy 301 GCCACATCGACCGTGAAGAGTGAACGCTTAAACTCCGATATTAAGTCCGCAATACGGCGGT 360
Db 241 GCTACTCTTGAATGAGTGAACGCGCAAAATTTGAAATGAGGTTTAAACAGTTCCGTTGAT 300
Qy 361 AATAACCGCGGCTGGTAAATGACAGCGCATGATTCAGCGTAAATGGTGGTCAAGTT 420
Db 301 GCGAAGCGTGTGAGTGAAGTTGACCAAGCTGACATCTCACTCCCTCCGTCAGTCACTGAGTT 360
Qy 421 GGTTTGGCAACAACGCGCAGCGGCTTAACCACTATTAA 456
Db 361 GGGTTTGGTAAACAACGCGCAGCGCTCACTCACTAA 396
```

RESULT 3

```
US-09-830-230A-629
Sequence 629, Application US/09830230A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
FILE OF INVENTION: Lyme Disease Vaccines
CURRENT APPLICATION NUMBER: US/09/830,230A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: PCT/US98/12718
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/057,483
PRIOR FILING DATE: 1997-09-03
PRIOR APPLICATION NUMBER: 60/053,344
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/053,377
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/050,359
PRIOR FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 756
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 629
LENGTH: 3990
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1135)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1148)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1210)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1244)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1247)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1250)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1251)
OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629
```

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.24;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTACGCTGGCTTCTCTGCAAGCGATGCCCTTAATCTGAACACCATTAACCG 216
Db 3433 GTTACGGCTGGCGCTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3492

Qy 217 AGCGGTATGTGTAACGCGCCGATTTAGGCCAGGTCGGATTAATGACTATTGAATG 276
Db 3493 ATTGCTGCTGTATTTGGAGAGGATATGAGATGCTGGGATTTTGTGAAGATGAGATG 3552

Qy 277 ACTGAGATGGTTTCAGAAATTAATGCAATTCAGACAGTGGAGCGTAAATCCGAT 336
Db 3553 AAGAGGATGATCAGATTGCTGCTCTATTGCTTTGAGGGGAGGCTTAAGAGTGAAG 3612

Qy 337 ATTAAGTGTGGCAATACGCGCGTAAATACG 367
Db 3613 TTTGCTGTGAAGAGATGATGATGAGAAAG 3643

RESULT 4
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antonie
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: GM-Negative EHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT FILING DATE: 2003-07-21
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Teiford, E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
; US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 2.8;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCGCAAGGATGCCCGTAAATCTGAACGACCATTAACCAAGCGGTTATGTAACG 232
Db 95342 CTGTGTAAACGCTGGCGCTTGTGTTGATGACGCTCATTTCCACGCAATGTTGCTTCA 95601

Qy 233 GCGCGCATGTAGCGGAGGTGCGGATATAGTACTATGAAGTCACTAGAGATGTTTCA 292
Db 95602 TCAAGCCCATATAGCCAGGACGAGATATGTTTCTGTATACCAATCTCTGGGCGCGTCA 95661

Qy 293 GAATTAATGCCACCATCAGCCAGTGG 318
Db 318

Db 95662 GAGAGCTTATTCACACGAGCTGTTG 95687

RESULT 5
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT FILING DATE: 2004-01-23
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 14522
; LENGTH: 1588
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
; US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.76; Indels 156; Gaps 0;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 40 GTTCTGCAAGTGTCTGCTGCGCGCTTTCACAAATGGGCGCGGCTAATCATTAAC 99
Db 455 GGTGTGGCGGCTTCGGTGGTGGTGTATGCGGCGCGGCTGATGTTGGTGGC 514

Qy 100 GCGCGCGCAATTAATTCGCGCGCGCACTCAACGTTAGCATTTATCAATACGTTCCGCT 159
Db 515 GCGCGTGGCTATGTTGT 574

Qy 160 AACGCTGCGTGTCTGCTGCAAGGATGCCCGTAAATCTGAAGAGACCATTAACCGAGC 219
Db 575 GGT 634

Qy 220 GGTATGTAAACGCGCGGATGTAGGCGGAGGTCGCGATTAATTAATTAATTAATTAATTA 279
Db 635 GGTATGTAAACGCTTGGCAGAGTAACTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 694

Qy 280 CAGATGTTTCAAAATTAATCCACCATGACCAAGTGA 320
Db 695 GCTGT 725

RESULT 6
US-10-788-792-27/c
; Sequence 27, Application US/10788792
; GENERAL INFORMATION:
; APPLICANT: Bayer Pharmaceuticals Corporation
; APPLICANT: Eveleigh, Deepa
; APPLICANT: Bigwood, Douglas
; TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
; FILE REFERENCE: 5152
; CURRENT FILING DATE: 2004-02-27
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,655
; PRIOR FILING DATE: 2003-02-28
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: Patent Version 3.2
; SEQ ID NO 27
; LENGTH: 1575
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.2;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

```

/ GENERAL INFORMATION: Application US/10767471
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CL001505 RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
/ CURRENT APPLICATION NUMBER: US/10/767,471
/ NUMBER OF SEQ ID NOS: 50231
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 10736
/ LENGTH: 24900
/ TYPE: DNA
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(24900)
/ OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Table
US-10-767-471-10736

```

```

Query Match          5.5%      Score 29.6/; DB 6; Length 24900;
Best Local Similarity 52.4%;   Pred. No. 11;
Matches 65; Conservative 0; Mismatches 59; Indels 0

```

Qy	207	CATTACCAAGAGCGTTATGTAAAGCGGCCGATGTAGCCAGGGTCCGATATATGTAC	266
Db	10048	CAGATCTCGATCGCTTTTATGTGTCTCAAAAGTCCTGGTCMAATGGTGCATATATTGGTG	9983
Qy	267	TATTGAACTGCACTCAGATGATGTTTGAAGATATATGACCATCGACCAAGTGAACGCTAA	326
Db	9988	TCTCGTAACCCAAACTGTATGTATGTAATAAATAATAGTGCATCTCTGTTTGTATATATTTAA	9929
Qy	327	AAAC 330	
Db	9928	AATC 9925	

RESULT 8
 US-10-767-701-28437
 ; Sequence 28437
 ; GENERAL INFORMATION:
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
 ; FILE REFERENCE: 38-21(53335)B
 ; CURRENT APPLICATION NUMBER: US/10/767,701
 ; CURRENT FILING DATE: 2004-01-29
 ; NUMBER OF SEQ. ID NOS: 63128
 ; SEQ. ID NO 28437
 ; LENGTH: 300
 ; TYPE: DNA
 ; ORGANISM: Sorghum bicolor
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: 7551668
 ; US-10-767-701-28437

Query Match	6.4%	Score 29.4	DB 6	Length 300
Best Local Similarity	54.1%	Prod No. 1.7		
Matches 60	Conservative	0	Mismatches 51	Indels 0
			Gaps	0

[illegible]

```

1 GENERAL INFORMATION:
2 APPLICANT: LADNER, ROBERT C.
3 APPLICANT: COHEN, EDWARD H.
4 APPLICANT: NASTRI, HORACIO G.
5 APPLICANT: ROOKEY, KRISTIN L.
6 APPLICANT: HOET, RENE
7 APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
8 TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
9 TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
10 TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
11 FILE REFERENCE: DVA/002 CIP2
12 CURRENT APPLICATION NUMBER: US/10/045,674A
13 CURRENT FILING DATE: 2001-10-25
14 PRIOR APPLICATION NUMBER: 06/198,069
15 PRIOR FILING DATE: 2000-04-17
16 PRIOR APPLICATION NUMBER: 09/837,306
17 PRIOR FILING DATE: 2001-04-17
18 NUMBER OF SEQ ID NOS: 635
19 SOFTWARE: PatentIn Ver. 2.1
20 SEQ ID NO 593
21 LENGTH: 1355
22 TYPE: DNA
23 ORGANISM: Artificial Sequence
24 FEATURE:
25 OTHER INFORMATION: Description of Artificial Sequence: M13-III
26 FEATURE:
27 NAME/KEY: CDS
28 LOCATION: (1)..(1305)
29 US-10-045-674A-593

```

Query Match	6.4%;	Score 29.4;	DB 6;	Length 1355;
Best Local Similarity	25.2%;	Pred. No. 3.4;		
Matches	39;	Conservative	50;	Mismatched

[illegible]

```

RESULT 10
US-60-545-213-2134/c
! Sequence 2134, Application US/60545213
! GENERAL INFORMATION:
! APPLICANT: Wyeth
! APPLICANT: Mounts, William Martin
! TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
! FILE REFERENCE: JMI01083 (031886-942099)
! CURRENT APPLICATION NUMBER: US/60/545,213
! CURRENT FILING DATE: 2004-02-18

```


NUMBER OF SEQ ID NOS: 303284
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2134
LENGTH: 1400
TYPE: DNA
ORGANISM: Homo sapiens
US-60-545-213-2134

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTGTAACGGCCGATGTAGCCAGGTCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGTTGCAATAGCGTGAACCGAGGTTGGGGTGGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATAATGCCACATGACAGTGAACGCTAAAACT 331
DB 377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGACATCTCTCACTCTGAACT 318
QY 332 CCGATATTACTGTGGCCCAATACGGCGGTAATAACGGCGCTGTTAATCAGACCGCAT 391
DB 317 CCGTTGAGCAGGCGGCTGATGCAATTAACCTGCTGATCTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCTAATGGTCGTCAGTTGGTTTGGCAACAAGCCACGGCTAACCG 450
DB 257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAGTAAGAGGGGGTGTACCTG 199

RESULT 11

US-60-545-213-2135/c
Sequence 2135, Application US/60545213
GENERAL INFORMATION:

APPLICANT: Mount, William Martin
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
FILE REFERENCE: AM101083 (031896-042099)
CURRENT APPLICATION NUMBER: US/60/545,213
CURRENT FILING DATE: 2004-02-18
NUMBER OF SEQ ID NOS: 303284
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2135
LENGTH: 1400
TYPE: DNA
ORGANISM: Homo sapiens
US-60-545-213-2135

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTGTAACGGCCGATGTAGCCAGGTCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGTTGCAATAGCGTGAACCGAGGTTGGGGTGGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATAATGCCACATGACAGTGAACGCTAAAACT 331
DB 377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGACATCTCTCACTCTGAACT 318
QY 332 CCGATATTACTGTGGCCCAATACGGCGGTAATAACGGCGCTGTTAATCAGACCGCAT 391
DB 317 CCGTTGAGCAGGCGGCTGATGCAATTAACCTGCTGATCTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCTAATGGTCGTCAGTTGGTTTGGCAACAAGCCACGGCTAACCG 450
DB 257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAGTAAGAGGGGGTGTACCTG 199

RESULT 12
US-60-545-213-6406/c
Sequence 6406, Application US/60545213
GENERAL INFORMATION:

APPLICANT: Mount, William Martin
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
FILE REFERENCE: AM101083 (031896-042099)
CURRENT APPLICATION NUMBER: US/60/545,213
CURRENT FILING DATE: 2004-02-18
NUMBER OF SEQ ID NOS: 303284
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6406
LENGTH: 1400
TYPE: DNA
ORGANISM: Homo sapiens
US-60-545-213-6406

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTGTAACGGCCGATGTAGCCAGGTCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGTTGCAATAGCGTGAACCGAGGTTGGGGTGGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATAATGCCACATGACAGTGAACGCTAAAACT 331
DB 377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGACATCTCTCACTCTGAACT 318
QY 332 CCGATATTACTGTGGCCCAATACGGCGGTAATAACGGCGCTGTTAATCAGACCGCAT 391
DB 317 CCGTTGAGCAGGCGGCTGATGCAATTAACCTGCTGATCTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCTAATGGTCGTCAGTTGGTTTGGCAACAAGCCACGGCTAACCG 450
DB 257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAGTAAGAGGGGGTGTACCTG 199

RESULT 13

US-60-545-213-6407/c
Sequence 6407, Application US/60545213
GENERAL INFORMATION:

APPLICANT: Mount, William Martin
TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
FILE REFERENCE: AM101083 (031896-042099)
CURRENT APPLICATION NUMBER: US/60/545,213
CURRENT FILING DATE: 2004-02-18
NUMBER OF SEQ ID NOS: 303284
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6407
LENGTH: 1400
TYPE: DNA
ORGANISM: Homo sapiens
US-60-545-213-6407

Query Match 6.4%; Score 29.4; DB 7; Length 1400;
Best Local Similarity 45.2%; Pred. No. 3.5;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTGTAACGGCCGATGTAGCCAGGTCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGTTGCAATAGCGTGAACCGAGGTTGGGGTGGATCGGTTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATAATGCCACATGACAGTGAACGCTAAAACT 331
DB 377 TTGGGGTTCTCCATCCACTGCTTGAAAGCCAGAGACATCTCTCACTCTGAACT 318
QY 332 CCGATATTACTGTGGCCCAATACGGCGGTAATAACGGCGCTGTTAATCAGACCGCAT 391
DB 317 CCGTTGAGCAGGCGGCTGATGCAATTAACCTGCTGATCTGTCATCAGGTTTCT 258
QY 392 CTGATTCACGCTAATGGTCGTCAGTTGGTTTGGCAACAAGCCACGGCTAACCG 450

Db 257 GGGATCTCCCTCCAGTTGTTGATATTCTTGAAGTGAATGAAGGGGTGTGACTTG 199

RESULT 14

US-10-771-241-117/c
; Sequence 117, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allen
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; PRIOR FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 117
; LENGTH: 249
; TYPE: DNA
; ORGANISM: E. Coli
US-10-771-241-117

Query Match

Best Local Similarity 6.4%; Score 29.2; DB 6; Length 249;
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 295 AATAATGCCACCATGACCAAGTGAACGCTAAACCTCCGATATTACTGTGCGCAATAC 354
Db 180 AAAATTGAAGCCATGACCTTTACCAAGCCAAACAGCGTGTGATCCATCCGCCGACTAC 121
QY 355 GCGGTATTAACGCGCGCTGTATTAACGCCGATCTGATTCAGCGTAAATGTCGT 414
Db 120 GGCACCGACTATCCCGACAGATAGTATTAAGATCCACTCATCTTACTGCGCAT 61
QY 415 CA 416
Db 60 GA 59

RESULT 15

US-10-771-241-29
; Sequence 29, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allen
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; PRIOR FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 757
; TYPE: DNA
; ORGANISM: E. Coli
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(757)
; OTHER INFORMATION: n = A,T,C or G
US-10-771-241-29

Query Match 6.4%; Score 29.2; DB 6; Length 757;
Best Local Similarity 52.5%; Pred. No. 3.1;
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 295 AATAATGCCACCATGACCAAGTGAACGCTAAACCTCCGATATTACTGTGCGCAATAC 354
Db 264 AAAATTGAAGCCATGACCTTTACCAAGCCAAACAGCGTGTGATCCATCCGCCGACTAC 323
QY 355 GCGGTATTAACGCGCGCTGTATTAACGCCGATCTGATTCAGCGTAAATGTCGT 414
Db 324 GGCACCGACTATCCCGACAGATAGTATTAAGATCCACTCATCTTACTGCGCAT 383
QY 415 CA 416
Db 384 GA 385

Search completed: March 17, 2004, 08:25:30
Job time : 71.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds

(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-2

Perfect score: 456

Sequence: 1 atgaataacaattgtctatc.....ctcggtcaccacgaagctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues 75154660

Total number of hits satisfying chosen parameters:
Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending Patents NA Main:*

- 1: /cgn2_6/ptodata/2/pna/PCRUS_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US106_COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
- 76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
- 77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
- 79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.*
- 80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.*
- 81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
- 82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
- 83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
- 84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
- 85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
- 86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
- 87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
- 88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
- 89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
- 90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
- 91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
- 92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
- 93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
- 94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
- 95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
- 96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
- 97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
- 98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
- 99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
- 100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
- 101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
- 102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
- 103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
- 104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*
- 105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq.*
- 106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq.*
- 107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq.*
- 108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq.*
- 109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq.*
- 110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-2	Sequence 2, Appl1
2	332.8	73.0	456	23	US-09-543-407-4	Sequence 4, Appl1
3	295	64.7	561	17	US-09-252-691-171	Sequence 171, Appl
4	295	64.7	561	17	US-09-252-691-C-171	Sequence 171, Appl
5	295	64.7	561	51	US-10-417-886-171	Sequence 171, Appl
6	74.4	16.3	100	45	US-10-146-442B-80	Sequence 80, Appl
7	66.4	14.6	100	45	US-10-146-442B-82	Sequence 82, Appl
8	64.8	14.2	100	45	US-10-146-442B-81	Sequence 81, Appl
9	44.2	9.7	438	31	US-09-739-449-3383	Sequence 3383, Appl
10	44.2	9.7	438	33	US-09-803-110-3383	Sequence 3383, Appl
11	44.2	9.7	29058	21	US-09-514-000-342	Sequence 342, Appl
12	44.2	9.7	187892	31	US-09-739-449-208	Sequence 208, Appl
13	44.2	9.7	187892	33	US-09-803-110-208	Sequence 208, Appl
14	41	9.0	7218	8	US-08-466-194-14	Sequence 14, Appl
15	37.6	8.2	456	23	US-09-543-407-13	Sequence 13, Appl
16	37.6	8.2	2000	36	US-09-887-272A-5263	Sequence 5263, Appl
17	36.6	8.0	500	55	US-10-764-432-554	Sequence 554, Appl
18	36	7.9	1230230	20	US-09-438-185-1	Sequence 1, Appl
19	36	7.9	1230230	20	US-09-438-185-1	Sequence 1, Appl
20	35.8	7.9	39	23	US-09-543-407-51	Sequence 51, Appl
21	35	7.7	738	50	US-10-369-435-3	Sequence 3, Appl1
22	35	7.7	105325	54	PCT-US03-38193-1574	Sequence 1574, Ap
23	35	7.7	105325	54	US-10-723-860-1574	Sequence 1574, Ap
24	34.6	7.6	1362	104	US-09-475-871-2480	Sequence 2480, A
25	34.4	7.5	2006	46	US-10-160-619-101	Sequence 101, Appl
26	34.4	7.5	18540	47	US-10-231-416-3	Sequence 3, Appl1
27	34.4	7.5	18540	47	US-10-231-426-3	Sequence 3, Appl1
28	34.4	7.5	20538	39	US-09-949-016-12887	Sequence 12887, A
29	34.4	7.5	20538	39	US-09-949-016-17052	Sequence 17052, A
30	34.4	7.5	432287	39	US-09-947-911-180	Sequence 180, App
31	34.4	7.5	322788	39	US-09-947-911-51	Sequence 51, App
32	34.2	7.5	456	23	US-09-543-407-25	Sequence 25, Appl
33	34.2	7.5	12757	36	US-09-897-516-2547	Sequence 2547, Ap
34	34.2	7.5	12757	36	US-09-897-516-2548	Sequence 2548, Ap
35	34.2	7.5	12757	36	US-09-897-516-2549	Sequence 2549, Ap
36	34.2	7.5	12757	36	US-09-897-516-2550	Sequence 2550, Ap
37	34.2	7.5	12757	36	US-09-897-516-2551	Sequence 2551, Ap
38	34.2	7.5	12757	36	US-09-897-516-2552	Sequence 2552, Ap
39	34.2	7.5	12757	36	US-09-897-516-2553	Sequence 2553, Ap
40	34.2	7.5	12757	36	US-09-897-516-2554	Sequence 2554, Ap
41	34.2	7.5	12757	36	US-09-897-516-2555	Sequence 2555, Ap
42	34.2	7.5	12757	36	US-09-897-516-2556	Sequence 2556, Ap
43	34.2	7.5	12757	36	US-09-897-516A-2547	Sequence 2547, Ap
44	34.2	7.5	12757	36	US-09-897-516A-2548	Sequence 2548, Ap
45	34.2	7.5	12757	36	US-09-897-516A-2549	Sequence 2549, Ap

ALIGNMENTS

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	Similarity	Score	DB	Length
Best Local Similarity	100.0%	100.0%	Pred. No. 10;	Indels 0; Gaps 0
Matches	456;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0
Query	1	ATGAAAAACAATGGTATTATTTATGATGTTGCAATACTGGGTGCGCTCGGATTCGAAC	60	
Db	1	ATGAAAAACAATGGTATTATTTATGATGTTGCAATACTGGGTGCGCTCGGATTCGAAC	60	
Qy	61	CGGCAAAATTATGATCTGCGCTGTTGAGAAATATTAATTTTGGCGTAAATGAATTAGCAAG	120	
Db	61	CGGCAAAATTATGATCTGCGCTGTTGAGAAATATTAATTTTGGCGTAAATGAATTAGCAAG	120	
Qy	121	TCTTCATTATTAATCAGCGCGCCATTAATGGTCAAGTCGCGACGGAATATAGTCCAGAGTA	180	
Db	121	TCTTCATTATTAATCAGCGCGCCATTAATGGTCAAGTCGCGACGGAATATAGTCCAGAGTA	180	
Qy	181	CGCCAGGAAGATCAAAACTATTTGTCGGTATTTTCAAGAAGAGGAATATATCGGCG	240	
Db	181	CGCCAGGAAGATCAAAACTATTTGTCGGTATTTTCAAGAAGAGGAATATATCGGCG	240	
Qy	241	AAATGCAACGAGGAGGAATTAATTAATCTTGGCTATTAATGAGCAAAACGGGCAATGCCAAC	300	
Db	241	AAATGCAACGAGGAGGAATTAATTAATCTTGGCTATTAATGAGCAAAACGGGCAATGCCAAC	300	
Qy	301	GATCCGAGTATATCGCAAGCGCTTACGGTATATAGTCAGCTATTATCAGAAAAGTTCT	360	
Db	301	GATCCGAGTATATCGCAAGCGCTTACGGTATATAGTCAGCTATTATCAGAAAAGTTCT	360	
Qy	361	GGAATTAAGGCGCAATATTAATCCAGTACGGTACGAGAAAACAGCAGTTGTAGTCAGAAA	420	
Db	361	GGAATTAAGGCGCAATATTAATCCAGTACGGTACGAGAAAACAGCAGTTGTAGTCAGAAA	420	
Qy	421	CAGTCGATATGCTATTCGGTCAACCAACGCTAA	456	
Db	421	CAGTCGATATGCTATTCGGTCAACCAACGCTAA	456	
RESULT 2				
US-09-543-407-4				
Sequence 4, Application US/09543407				
GENERAL INFORMATION:				
APPLICANT: White, Aaron P.				
APPLICANT: Dotan, James L.				
APPLICANT: Collinson, S. Karen				
APPLICANT: Kay, William W.				
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR				
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES				
FILE REFERENCE: 920943.406				
CURRENT APPLICATION NUMBER: US/09/543.407				
CURRENT FILING DATE: 2000-04-05				
NUMBER OF SEQ ID NOS: 59				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 4				
LENGTH: 456				
TYPE: DNA				
ORGANISM: E. Coli				
US-09-543-407-4				
Query Match	73.0%	Score 332.8;	DB 23;	Length 456;
Best Local Similarity	83.1%;	Pred. No. 7.2e-92;		
Matches	379;	Conservative	0;	Mismatches 77; Indels 0; Gaps 0;
Qy	1	ATGAAAAACAATGGTATTATTTATGATGTTGCAATACTGGGTGCGCTCGGATTCGAAC	60	
Db	1	ATGAAAAACAATGGTATTATTTATGATGTTGCAATACTGGGTGCGCTCGGATTCGAAC	60	
Qy	61	CGGCAAAATTATGATCTGCGCTGTTGAGAAATATTAATTTTGGCGTAAATGAATTAGCAAG	120	
Db	61	CGGCAAAATTATGATCTGCGCTGTTGAGAAATATTAATTTTGGCGTAAATGAATTAGCAAG	120	
Qy	121	TCTTCATTATTAATCAGCGCGCCATTAATGGTCAAGTCGCGACGGAATATAGTCCAGAGTA	180	
Db	121	TCTTCATTATTAATCAGCGCGCCATTAATGGTCAAGTCGCGACGGAATATAGTCCAGAGTA	180	
Qy	181	CGCCAGGAAGATCAAAACTATTTGTCGGTATTTTCAAGAAGAGGAATATATCGGCG	240	
Db	181	CGCCAGGAAGATCAAAACTATTTGTCGGTATTTTCAAGAAGAGGAATATATCGGCG	240	
Qy	241	AAATGCAACGAGGAGGAATTAATTAATCTTGGCTATTAATGAGCAAAACGGGCAATGCCAAC	300	
Db	241	AAATGCAACGAGGAGGAATTAATTAATCTTGGCTATTAATGAGCAAAACGGGCAATGCCAAC	300	
Qy	301	GATCCGAGTATATCGCAAGCGCTTACGGTATATAGTCAGCTATTATCAGAAAAGTTCT	360	
Db	301	GATCCGAGTATATCGCAAGCGCTTACGGTATATAGTCAGCTATTATCAGAAAAGTTCT	360	
Qy	361	GGAATTAAGGCGCAATATTAATCCAGTACGGTACGAGAAAACAGCAGTTGTAGTCAGAAA	420	
Db	361	GGAATTAAGGCGCAATATTAATCCAGTACGGTACGAGAAAACAGCAGTTGTAGTCAGAAA	420	
Qy	421	CAGTCGATATGCTATTCGGTCAACCAACGCTAA	456	
Db	421	CAGTCGATATGCTATTCGGTCAACCAACGCTAA	456	

Db 121 TCTTATTATACAGGAGCCATATGCTGACGCTGGAGCAATATATATGCTCAGTTA 180
Qy 181 CGCCAGGAAGATCAAACTATGTCCTGTTATTTACAGAGAGAAATATGCGGCG 240
Db 181 CGCCAGGAGGCTCAAACTTTGGCGGTGTTGGCAGAAAGATGATGCAACCGGCA 240
Qy 241 AAAGTCGACGAGGAGGAAATATATCTTGGCTATATTTAGCAAAACGGGCAATGCCAAC 300
Db 241 AAGATTGACACAGAGAGATATATACCTTGATATATTTAGTCAGCGCGGCTGCCAAC 300
Qy 301 GATGCCAGTATATCGCAAGCGCTTACGCTATAGTGCAGCTATTATCCAGAAAGTTCT 360
Db 301 GATGCCAGTATTTCCCAAGTCTTATGTTATATCTGCAATATTTATCCAGAAAGTTCT 360
Qy 361 GGAATTAAGCCCAATATTTACCGATACGCTACGCAAGAAACAGCAGTTGATGCAAAA 420
Db 361 GGTATTAAGCAAAATTTACACAGATGCTCAAAAAACGGCAATGTTAGTGCAGAGA 420
Qy 421 CAGTCGATATGCTATTCGGCTACCAACGCTAA 456
Db 421 CAGTCGCAAAATGCTATTCGGCTGAGACACAGTTAA 456

RESULT 3
US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; PRIORITY FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-171

Query Match 64.7%; Score 295; DB 17; Length 561;
Best Local Similarity 78.0%; Pred. No. 3.9e-80;
Matches 355; Conservative 0; Mismatches 100; Indels 0; Gaps 0;
Qy 1 ATGAAAAACAATTTGTTATTTATGATGTTGACAATACCTGGGCTGGGATTCACAAC 60
Db 106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGGCTGGTTGTAATC 165
Qy 61 GCGACAATTTATGATCTGGCTGCTCAGAAATATTTTGGCGTAAATGAATTAAGCAAG 120
Db 166 GCGAGTATTCAGATTAGCCAGTCTGAAATATATTTTGGCATATTAATGAATTAAGTAA 225
Qy 121 TCTTCATTTAATCAGCGCGCCATTTATGTCAGTCGCGACGAGTAATATGTCAGAGTA 180
Db 226 GCTTATACATCAGCGCCATTTATGTCAGCGGTTCAGAAATATTTCTGATGTA 285
Qy 181 CGCCAGGAAGATCAAACTATGTCCTTATTTCAAGAAAGAGAAATATTCGGCGC 240
Db 286 CGCCAGGACGCTTCAAAATGCTGCTGCTATTTCTCAGAGGCGGGAATTAACGCGCG 345
Qy 241 AAAGTCGACGAGGAGGAAATATATCTTGGCTATATTTAGCAAAACGGGCAATGCCAAC 300
Db 346 AATGTTATCAGTCAGGAGCATATACCTTGCTATATATGATCAACCGCAACGCGAAC 405
Qy 301 GATGCCAGTATATCGCAAGCGCTTACGCTATAGTGCAGCTATTATCCAGAAAGTTCT 360
Db 406 GATGCCAGTATTAACGAGGCGCTTTGGCAACACCGCATGATATATCCAGAAAGCTCG 465
Qy 361 GGAATTAAGCCCAATATTTACCGATACGCTACGCAAGAAACAGCAGTTGATGCAAAA 420
Db 466 GGTATTAAGCAAAATTTACACAGATGCTCAAAAAACGGCAATGTTAGTGCAGAGA 525

Qy 421 CAGTCGATATGCTATTCGGCTACCCAGGCTA 455
Db 526 CAGTCGCAAAATGCTATTCGGCTATTCAGGTTA 560

RESULT 4
US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; PRIORITY FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIORITY FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-171

Query Match 64.7%; Score 295; DB 17; Length 561;
Best Local Similarity 78.0%; Pred. No. 3.9e-80;
Matches 355; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

Qy 1 ATGAAAAACAATTTGTTATTTATGATGTTGACAATACCTGGGCTGGGATTCACAAC 60
Db 106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGGCTGGTTGTAATC 165
Qy 61 GCGACAATTTATGATCTGGCTGCTCAGAAATATTTTGGCGTAAATGAATTAAGCAAG 120
Db 166 GCGAGTATTCAGATTAGCCAGTCTGAAATATATTTTGGCATATTAATGAATTAAGTAA 225
Qy 121 TCTTCATTTAATCAGCGCGCCATTTATGTCAGTCGCGACGAGTAATATGTCAGAGTA 180
Db 226 GCTTATACATCAGCGCCATTTATGTCAGCGGTTCAGAAATATTTCTGATGTA 285
Qy 181 CGCCAGGAAGATCAAACTATGTCCTTATTTCAAGAAAGAGAAATATTCGGCGC 240
Db 286 CGCCAGGACGCTTCAAAATGCTGCTGCTATTTCTCAGAGGCGGGAATTAACGCGCG 345
Qy 241 AAAGTCGACGAGGAGGAAATATATCTTGGCTATATTTAGCAAAACGGGCAATGCCAAC 300
Db 346 AATGTTATCAGTCAGGAGCATATACCTTGCTATATATGATCAACCGCAACGCGAAC 405
Qy 301 GATGCCAGTATATCGCAAGCGCTTACGCTATAGTGCAGCTATTATCCAGAAAGTTCT 360
Db 406 GATGCCAGTATTAACGAGGCGCTTTGGCAACACCGCATGATATATCCAGAAAGCTCG 465
Qy 361 GGAATTAAGCCCAATATTTACCGATACGCTACGCAAGAAACAGCAGTTGATGCAAAA 420
Db 466 GGTATTAAGCAAAATTTACACAGATGCTCAAAAAACGGCAATGTTAGTGCAGAGA 525
Qy 421 CAGTCGATATGCTATTCGGCTACCCAGGCTA 455
Db 526 CAGTCGCAAAATGCTATTCGGCTATTCAGGTTA 560

RESULT 5
US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417,886
; PRIORITY FILING DATE: 2003-04-17

Query Match 16.3%; Score 74.4; DR 45; Length 100.

Query Match	14.2%	Score 64.8;	DB 45;	Length 100;
Best Local Similarity	78.0%	Pred. No. 4.4e-09;		
Matches	78;	Conservative	0.0	Matches

RESULT 12
US-09-739-449-208
; Sequence 208, Application US/09739449

GENERAL INFORMATION:
 APPLICANT: Hinkle, Gregory J.
 TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 FILE REFERENCE: 38-10(15490) C
 CURRENT APPLICATION NUMBER: US/09/739,449
 PRIOR FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 09/514,000
 NUMBER OF SEQ ID NOS: 13351
 SEQ ID NO 208
 LENGTH: 187892
 TYPE: DNA
 ORGANISM: Agrobacterium tumefaciens
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(187892)
 OTHER INFORMATION: unsure at all n locations
 US-09-739-449-208

Query Match 9.7%; Score 44.2; DB 31; Length 187892;
 Best Local Similarity 47.6%; Pred. No. 0.14;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

QY 145 ATTGTCAGTCCGACGATTAATAGTCCAGAGTACGCGAGAAAGATCAAACTATTG 204
 DB 137108 ATCGAACATATGCTGTCTCAACTCTGCCGGGGGCGCCAGAAAGGTATGGCAACCGA 137167
 QY 205 TCCGTTATTTCACAGAGAGAGAAATATTCGGGCGAAAGTGCACGAGGAGATTTAT 264
 DB 137168 ATCCGACCTACCAAGACGCGGTTTCAACCGGATCGTCGGCCACAGTATGTCTCAC 137227
 QY 265 AACTTGCCTATTATTGAGCAAAACGCGCATGCGCAACATGCGCAATATTCGCAAGGCT 324
 DB 137228 AACCTTCCGCTGTGCTGAGAGAGGAGATGACATTCAGTCCACACCAAGAGCGC 137287
 QY 325 TACGGTAATATGACGCTATTATCCAGAAAGTTTGGAAATTAAGGCCAATATTTACCAG 384
 DB 137288 AACCGCAAGTGGCGGCGATCGCGCATTCGCTCCAAACACACCATCTGACCCAG 137347
 QY 385 TACGGTACGCAAAACAGCAGTTTGTAGTCAG 417
 DB 137348 GACGGCAATGGCAACATTCGTGACAGCGTTTCA 137380

RESULT 13
 US-09-803-110-208
 Sequence 208, Application US/09803110
 GENERAL INFORMATION:
 APPLICANT: Hinkle, Gregory J.
 TITLE OF INVENTION: Agrobacterium tumefaciens Genome Sequences and Uses Thereof
 FILE REFERENCE: 38-10(15490) D
 CURRENT APPLICATION NUMBER: US/09/803,110
 PRIOR FILING DATE: 2001-03-12
 PRIOR APPLICATION NUMBER: US 09/739,449
 PRIOR FILING DATE: 2000-12-19
 PRIOR APPLICATION NUMBER: US 09/514,000
 PRIOR FILING DATE: 2000-02-23
 PRIOR APPLICATION NUMBER: US 60/168,139
 NUMBER OF SEQ ID NOS: 13351
 SEQ ID NO 208
 LENGTH: 187892
 TYPE: DNA
 ORGANISM: Agrobacterium tumefaciens
 FEATURE:
 NAME/KEY: unsure
 LOCATION: (1)..(187892)
 OTHER INFORMATION: unsure at all n locations
 US-09-803-110-208

Query Match 9.7%; Score 44.2; DB 33; Length 187892;

Best Local Similarity 47.6%; Pred. No. 0.14;
 Matches 130; Conservative 0; Mismatches 143; Indels 0; Gaps 0;
 QY 145 ATTGTCAGTCCGACGATTAATAGTCCAGAGTACGCGAGAAAGATCAAACTATTG 204
 DB 137108 ATCGAACATATGCTGTCTCAACTCTGCCGGGGGCGCCAGAAAGGTATGGCAACCGA 137167
 QY 205 TCCGTTATTTCACAGAGAGAGAAATATTCGGGCGAAAGTGCACGAGGAGATTTAT 264
 DB 137168 ATCCGACCTACCAAGACGCGGTTTCAACCGGATCGTCGGCCACAGTATGTCTCAC 137227
 QY 265 AACTTGCCTATTATTGAGCAAAACGCGCATGCGCAACATGCGCAATATTCGCAAGGCT 324
 DB 137228 AACCTTCCGCTGTGCTGAGAGAGGATGACATTTAGTTCCACCAACGAGAGCGC 137287
 QY 325 TACGGTAATATGACGCTATTATTCAGAAAGTTTGGAAATTAAGGCCAATATTTACCAG 384
 DB 137288 AACCGCAAGTGGCGGCGATCGCGCATTCGCTCCAAACACACCATCTGACCCAG 137347
 QY 385 TACGGTACGCAAAACAGCAGTTTGTAGTCAG 417
 DB 137348 GACGGCAATGGCAACATTCGTGACAGCGTTTCA 137380

RESULT 14
 US-08-466-194-14/c
 Sequence 14, Application US/08466194
 GENERAL INFORMATION:
 APPLICANT: DORNER, F.
 APPLICANT: SCHEITLINGER, F.
 APPLICANT: FALKNER, F. G.
 TITLE OF INVENTION: RECOMBINANT FOXP2 VIRUS
 NUMBER OF SEQUENCES: 52
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, NW, Suite 500
 CITY: Washington
 STATE: DC
 COUNTRY: USA
 ZIP: 20007
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,194
 FILING DATE: 06-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/232,463
 FILING DATE: 22-APR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/935,313
 FILING DATE: 26-AUG-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP 91 114 300.6
 FILING DATE: 26-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 30472/201 IMMU
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 672-5300
 TELEFAX: (202) 672-5399
 TELEX: 904136
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 7218 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 CLONE: pTZgpt-F15

US-08-466-194-14

Query Match 9.0%; Score 41; DB 8; Length 7218;

Best Local Similarity 3.0%; Pred. No. 0.44; Matches 11; Conservative 200; Mismatches 150; Indels 0; Gaps 0;

OY 63 GACAAATATGATCGCTGCTTCAATATATTTGCGGAATGATTAAGCAAGTC 122
 DB 1454 GAGATGAAGAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1395
 OY 123 TTCAATTAATGACGGCCATTATGTCAGTCGCAAGATTAATGTCAGATAG 182
 DB 1394 RRR 1335
 OY 183 CCAGAAAGATCAAAACTATTGTCCTATTATTCACAGAGAGAAATATCGGCCA 242
 DB 1334 RRR 1275
 OY 243 AGTCAGCAGGAGGAAATTAATCTTGCATATATGAGCAAGGCAATGCCAAGA 302
 DB 1274 RRR 1215
 OY 303 TGCCAGTATATCGCAAGCGCTTACGTAATAGTCAGCTATTATCAGAAAGTTCTG 362
 DB 1214 RRR 1155
 OY 363 AAATAGGCCAATATATCCAGTAGCGTAAGCAAGAAACAGCAGTTGATGCAAGACA 422
 DB 1154 RRR 1095
 OY 423 G 423
 DB 1094 R 1094

RESULT 15

US-09-543-407-13

; Sequence 13, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Dorian, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL, FIMBRIAL, SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-13

Query Match 8.2%; Score 37.6; DB 23; Length 456;

Best Local Similarity 51.2%; Pred. No. 2; Matches 88; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

OY 272 CGTATATTAGCAAAAGGCAATGCAAGATGCAATATTCGCAAGCGCTTACGCTA 331
 DB 203 CGACCATTAACCAAGCGGTTATGTTAAGCGCGCATGTAAGCCAGGTCGGATATA 262
 OY 332 ATAGTCAGCTATTATTCAGAAAGTTCTGGAATAAGCCCAATATTACCAATACGTA 391
 DB 263 GTACTATTGAAGTCTAGATGTTTCAGAAATTAATGCAACATGCAACGATGGAACG 322
 OY 392 CGCAGAAAAAGCAGTTGTAGTGCAGAAAGTCGATATGCTATTGCGCT 443

DB 323 CTAAAACTCGATATTACTGTGCGCCAAATGATCACTGTTACCGCTGT 374

Search completed: March 18, 2004, 02:46:45
 Job time : 2935.24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using bw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-2

Perfect Score: 456
Sequence: 1 atgaaacaatgttatt.....ttcgctcaccacgctaa 456

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1: Pending Parents NA New
2: /cgn2_6/pilotdata/2/pna/PCT_NEW_COMB.seq.*
3: /cgn2_6/pilotdata/2/pna/US06_NEW_COMB.seq.*
4: /cgn2_6/pilotdata/2/pna/US07_NEW_COMB.seq.*
5: /cgn2_6/pilotdata/2/pna/US08_NEW_COMB.seq.*
6: /cgn2_6/pilotdata/2/pna/US09_NEW_COMB.seq.*
7: /cgn2_6/pilotdata/2/pna/US10_NEW_COMB.seq.*
8: /cgn2_6/pilotdata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	36.6	8.0	500	1	PCT-US04-02000-554	Sequence 554, Apple
2	33.8	7.4	40143	6	US-10-765-790-79	Sequence 79, Apple
3	32.8	7.2	418	6	US-10-779-543-9321	Sequence 9321, Apple
4	32.4	7.1	201	6	US-10-767-471-8835	Sequence 8835, Apple
5	32.4	7.1	201	6	US-10-767-471-34493	Sequence 34493, Apple
6	32.4	7.1	1227	6	US-10-767-471-501	Sequence 501, Apple
7	32.4	7.1	76573	6	US-10-767-471-10718	Sequence 10718, Apple
8	32	7.0	1310	6	US-10-767-795-2887	Sequence 2887, Apple
9	31.8	7.0	5495	6	US-10-765-169-342	Sequence 342, Apple
10	31.2	6.8	201	6	US-10-767-471-8837	Sequence 8837, Apple
11	31.2	6.8	201	6	US-10-767-471-54439	Sequence 54439, Apple
12	31	6.8	394468	7	US-60-548-091-5725	Sequence 5725, Apple
13	30.8	6.8	309	6	US-10-257-047-43	Sequence 43, Apple
14	30.8	6.8	309	6	US-10-257-047-46	Sequence 46, Apple
15	30.8	6.8	3147	6	US-10-257-047-44	Sequence 44, Apple
16	30.8	6.8	3147	6	US-10-257-047-45	Sequence 45, Apple
17	30.8	6.8	3147	6	US-10-257-047-47	Sequence 47, Apple
18	30.8	6.8	3147	6	US-10-257-047-48	Sequence 48, Apple
19	29.8	6.5	592	6	US-10-767-701-25359	Sequence 25359, Apple
20	29.8	6.5	1035	6	US-10-779-543-7983	Sequence 7983, Apple
21	29.8	6.5	2583	6	US-10-441-926-25	Sequence 25, Apple
22	29.8	6.5	2583	6	US-10-441-926-27	Sequence 27, Apple
23	29.8	6.5	2583	6	US-10-441-926-29	Sequence 29, Apple
24	29.6	6.5	1084	6	US-10-487-578-48	Sequence 48, Apple
25	29.6	6.5	1084	6	US-10-487-578-33	Sequence 33, Apple
26	29.6	6.5	1217	6	US-10-487-578-46	Sequence 46, Apple

ALIGNMENTS

27	29.6	6.5	1219	6	US-10-487-578-44	Sequence 44, Appl
28	29.4	6.4	782	6	US-10-767-701-3762	Sequence 3762, Ap
C	29.4	6.4	188053	6	US-10-417-375A-125	Sequence 125, App
30	29.2	6.4	344805	6	US-10-779-271-1	Sequence 1, Appl
31	29.2	6.4	354592	6	US-10-765-790-70	Sequence 70, Appl
32	29	6.4	1287	6	US-10-100-683-1253	Sequence 1253, Ap
C	29	6.4	1382	6	US-10-100-683-3753	Sequence 3753, Ap
C	29	6.4	113486	6	US-10-417-375A-18	Sequence 18, Appl
33	29	6.4	156652	6	US-10-765-790-48	Sequence 48, Appl
34	29	6.4	156652	6	US-10-767-471-10755	Sequence 10755, A
C	29	6.3	19959	6	US-10-767-471-10892	Sequence 10892, A
35	28.8	6.3	53118	6	US-10-779-543-12419	Sequence 12419, A
36	28.8	6.3	346	6	US-09-741-873C-3	Sequence 3, Appl
C	28.6	6.3	675	5	US-10-417-884A-2080	Sequence 2080, Ap
37	28.6	6.3	1743	6	US-10-486-977-58	Sequence 58, Appl
38	28.6	6.3	3537	6	US-10-417-375A-63	Sequence 63, Appl
C	41	28.6	219715	6	US-10-771-241-139	Sequence 139, App
41	28.6	6.2	1149	6	US-60-548-091-9	Sequence 9, Appl
42	28.4	6.2	2525	7	US-60-548-091-8	Sequence 8, Appl
43	28.4	6.2	2665	7		
44	28.4	6.2				
45	28.4	6.2				

RESULT 1

```

PCT-US04-02000-554
Sequence 554, Application PC/TUS0402000
GENERAL INFORMATION:
APPLICANT: Science Applications International Corporation
TITLE OF INVENTION: Method and System for Identifying Biological Entities in
TITLE OF INVENTION: Biological and Environmental Samples
FILE REFERENCE: 36609-183264 (SAIC0087-PCT)
CURRENT APPLICATION NUMBER: PCT/US04/02000
CURRENT FILING DATE: 2004-01-23
PRIOR APPLICATION NUMBER: US 60/441,745
PRIOR FILING DATE: 2003-01-23
PRIOR APPLICATION NUMBER: US 60/441,806
PRIOR FILING DATE: 2003-01-23
NUMBER OF SEQ ID NOS: 3242
SOFTWARE: PatentIn version 3.2
SEQ ID NO 554
LENGTH: 500
TYPE: DNA
ORGANISM: Clostridium perfringens
PCT-US04-02000-554

```

Query Match	8.0%: Score 36.6; DB 1; Length 500;
Best Local Similarity	44.4%; Pred. No. 0.051;
Matches 147; Conservative	0; Mismatches 184; Indels 0; Gaps 0
QY	57 AACGCCACAATATGATCTGCCTCGTTCAGATATATATTTGCGTAAATGAAATTAG 116
Db	68 AAATCTGTGAATTTAAAGTAAATCTCGTGAAGAGTTATGAAGTTATTAATTAATATA 127
QY	117 CAACTCTTCATTTAATCAGCGCGCCATTATTTGTCAGATCGGCACGGATTAATAGTCAG 176
Db	128 CTAACTACTAGTCTATTGCTAGAAAGTCCTTCAAAAAGTTATTAATACTAATCTATTGAGT 187
QY	177 AGTACGCCAGAAAGATTCAAAACATATGTCCGTATTTCACAAAGAGAGAAATTAATCG 236
Db	188 TATAGTTCCTGATATTTACAAATCCATCTTTGGATCAATATATTAAGGAATTAAGATGT 247
QY	237 GGCAGAAATGCACGACGAGGAGATTTAATCTTGGCGTATTTGACAAACGGCGCAATGC 296
Db	248 TGCAGAAATGCTAATCTTAATTTGATTCCTTTGAGCTCTACAGAGATATCGATAGGGA 307
QY	297 CAACGATGCCAGTATATCGCAAGCGCTTACGTTAATAGTGCACGACTTATATCCAGAAAG 356
Db	308 AATTAAGGCTATTAACATTTAAAGAAACAAAGAAATTAGGGGAATTAATATTTGCTAAC 367
QY	357 TTCTGGAATTAAGCCCAATATTAACCGATAC 387
Db	368 CTCGTGTGAATAATGATTTAAATACCGATAC 398

RESULT 2

US-10-765-790-79/c
Sequence 79, Application US/10765790
GENERAL INFORMATION:
APPLICANT: Bayer Healthcare LLC
APPLICANT: Beard, Chris
APPLICANT: Burgess, Chris
APPLICANT: Gannon, Allison
APPLICANT: Harvey, Jeanne
APPLICANT: Lechner, John F.
APPLICANT: Li, Zheng
TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
FILE REFERENCE: 1657/2035
CURRENT APPLICATION NUMBER: US/10/765,790
CURRENT FILING DATE: 2004-01-27
PRIOR APPLICATION NUMBER: US 10/737,082
PRIOR FILING DATE: 2003-12-16
NUMBER OF SEQ ID NOS: 300
SOFTWARE: PatentIn version 3.2
SEQ ID NO 79
LENGTH: 401433
TYPE: DNA
ORGANISM: Homo sapiens
US-10-765-790-79

Query Match
Best Local Similarity 7.4%; Score 33.8; DB 6; Length 401433;
Matches 83; Conservative 0; Mismatches 82; Indels 0; Gaps 0;

QY 5 AAACAATGTTATTTATATATGATGACAACTGGGCTGGGATGCAACCGGA 64
DB 270713 AGAAATTTTCTTTCTTTTGTGATGTTCTGAACTCGGAACAACATGT 270654
QY 65 CAATTTATGATCTGCTGCTGATATATATTTGGGTAATGATTAAGCAAGTCTT 124
DB 270653 CAATGTAATGATGCTTCCCTCGAATAATGTTTGAATAATTAACCAACAAT 270594
QY 125 CATTATACAGGGGCGCATTTATGTCAGAGTCGACGAGATTA 169
DB 270593 ACTAAGAACAGGAACACAGATGAGTATGTGAAACATAGAAAT 270549

RESULT 3

US-10-779-543-9321
Sequence 9321, Application US/10779543
GENERAL INFORMATION:
APPLICANT: Williams et al
TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
FILE REFERENCE: 2300-21302
CURRENT APPLICATION NUMBER: US/10/779,543
CURRENT FILING DATE: 2004-02-12
PRIOR APPLICATION NUMBER: 10/076,555
PRIOR FILING DATE: 2002-02-15
PRIOR APPLICATION NUMBER: 09/217,471
PRIOR FILING DATE: 1998-12-21
PRIOR APPLICATION NUMBER: 60/068,755
PRIOR FILING DATE: 1997-12-23
PRIOR APPLICATION NUMBER: 60/080,664
PRIOR FILING DATE: 1998-04-03
PRIOR APPLICATION NUMBER: 60/105,234
PRIOR FILING DATE: 1998-10-21
PRIOR APPLICATION NUMBER: 09/297,648
PRIOR FILING DATE: 2000-04-10
PRIOR APPLICATION NUMBER: PCT/US99/01619
PRIOR FILING DATE: 1999-01-28
PRIOR APPLICATION NUMBER: 60/072,910
PRIOR FILING DATE: 1998-01-28
PRIOR APPLICATION NUMBER: 60/075,954
PRIOR FILING DATE: 1998-02-24
PRIOR APPLICATION NUMBER: 60/080,114

PRIOR FILING DATE: 1998-03-31

Remaining Prior Application data removed - See File Wrapper or PALM.

NUMBER OF SEQ ID NOS: 23767
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9321

LENGTH: 418
TYPE: DNA
ORGANISM: Homo sapiens
US-10-779-543-9321

Query Match
Best Local Similarity 7.2%; Score 32.8; DB 6; Length 418;
Matches 70; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 87 AGAATTAATTTTGGCGTAATGATTAAGCAAGTCTTATTTATTCAGCGCCATTAT 146
DB 208 AGAATTCAGACTGGGGCCATGACCAAGGCAATCTGACTCAAGATCTTGA 267
QY 147 TGGTCAAGTGGGACCGGATTAATGATCCAGAGTACGCCAGAGGATCAAACTATTGTC 206
DB 268 AGATTGATTAGCAGTGTGCTTGAAACAGACTGGGCCAGGAGGCTATTAAGCAGGAG 327
QY 207 CGTTATTGACA 218
DB 328 AGTCATTAA 339

RESULT 4

US-10-767-471-8836/c
Sequence 8836, Application US/10767471
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8836
LENGTH: 201
TYPE: DNA
ORGANISM: Homo sapiens
US-10-767-471-8836

Query Match
Best Local Similarity 7.1%; Score 32.4; DB 6; Length 201;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

QY 326 ACGGTATAGTGAAGTATTTATCCAGAAAGTTTGGAAATTAAGCCAAATTAACCCAGT 385
DB 142 ATGGAGATATGAGGGGAGAAATCCAGATGATGATTAATTAAGGAGCTATTTCCAGG 83
QY 386 ACGGTACGAGAAACAGCAGTGTGATGACAGAAACAGTCGATATGG 433
DB 82 AGATTATCATGTATCATGATGAGGACAGACAGAAACAGAAAGCTG 35

RESULT 5

US-10-767-471-34493/c
Sequence 34493, Application US/10767471
GENERAL INFORMATION:
APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 34493
LENGTH: 201
TYPE: DNA

ORGANISM: Homo sapiens
US-10-767-471-34493

Query Match 7.1%; Score 32.4; DB 6; Length 201;
Best Local Similarity 55.6%; Pred. No. 0.78;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCAATATTACCCAGT 385
Db 142 ATGGAGATGAGGCGGAGAAATCCAAAGTGTGATGATTAATPAGAGGCTATTTCACAG 83

Qy 386 ACGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCCGATATG 433
Db 82 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 35

RESULT 6
US-10-767-471-501/c
Sequence 501, Application US/10767471
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 501
LENGTH: 1227
TYPE: DNA
ORGANISM: Homo sapiens
US-10-767-471-501

Query Match 7.1%; Score 32.4; DB 6; Length 1227;
Best Local Similarity 55.6%; Pred. No. 1.3;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCAATATTACCCAGT 385
Db 476 ATGGAGATGAGGCGGAGAAATCCAAAGTGTGATGATTAATPAGAGGCTATTTCACAG 417

Qy 386 ACGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCCGATATG 433
Db 416 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 369

RESULT 7
US-10-767-471-10718/c
Sequence 10718, Application US/10767471
GENERAL INFORMATION:

APPLICANT: CARGILL, Michele et al.
TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
FILE REFERENCE: CLO01505
CURRENT APPLICATION NUMBER: US/10/767,471
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 10718
LENGTH: 76573
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(76573)
OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10718

Query Match 7.1%; Score 32.4; DB 6; Length 76573;
Best Local Similarity 55.6%; Pred. No. 4.3;
Matches 60; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

Qy 326 ACGTAATAGTGCAGCTATTATCCAGAAAGTTCTGAAATPAGGCCAATATTACCCAGT 385
Db 69821 ATGGAGATGAGGCGGAGAAATCCAAAGTGTGATGATTAATPAGAGGCTATTTCACAG 69762
Qy 386 ACGTACCCAGAAAAACAGCAGTTGTAGTCAGAAAAAGTCCGATATG 433
Db 69761 AGGATTATCATGTACATGATGAGCAGAGCAGAAACAGAAAAAGCTGG 69714

RESULT 8
US-10-767-795-2887
Sequence 2887, Application US/10767795
GENERAL INFORMATION:

APPLICANT: KOVALLIC, David K.
APPLICANT: Cao, Yongwei
APPLICANT: Zhou, Yihua
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21(53534)B
CURRENT APPLICATION NUMBER: US/10/767,795
CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 117596
SEQ ID NO 2887
LENGTH: 1310
TYPE: DNA
ORGANISM: Gossypium hirsutum
FEATURE:
OTHER INFORMATION: clone ID: GOSHI-09MAY01-C71322_1
US-10-767-795-2887

Query Match 7.0%; Score 32; DB 6; Length 1310;
Best Local Similarity 55.4%; Pred. No. 1.8;
Matches 62; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Qy 2 TGAACAAATTTGTTATTTATGATGTTGACAAATATCTGGTCCCTGGGATTGCCAACCG 61
Db 626 TGATGATTCATTTCTTCCCATGAGATGGAGACATTGCCGAAATATGATGATGCGG 685

Qy 62 CGACAAATTTATGATCTGGCTCTGTCAGAAATTAATTTGGCGTAATGAATT 113
Db 686 TGTCAAAATTCAGACCTGCCAATCAACGTAACAAATGCTCAAAATGAAGT 737

RESULT 9
US-10-775-169-342/c
Sequence 342, Application US/10775169
GENERAL INFORMATION:

APPLICANT: WYETH
APPLICANT: Burczynski, Michael
APPLICANT: Twine, Natalie
APPLICANT: Dornier, Andrew
TITLE OF INVENTION: Method for Monitoring Drug Activities in Vivo
FILE REFERENCE: AM101080 (031896-013000)
CURRENT APPLICATION NUMBER: US/10/775,169
CURRENT FILING DATE: 2004-02-11
NUMBER OF SEQ ID NOS: 5278
SOFTWARE: PatentIn version 3.2
SEQ ID NO 342
LENGTH: 5495
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-169-342

Query Match 7.0%; Score 31.8; DB 6; Length 5495;
Best Local Similarity 51.0%; Pred. No. 3.1;
Matches 75; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

Qy 2 TGAACAAATTTGTTATTTATGATGTTGACAAATCTGGTCCCTGGGATTGCCAACCG 61
Db 4776 TGCAAAACAGAGATTTCTCGTTTGAAAAAACAAGCAAGAAACAGGTGATATCA 4717
Qy 62 CGACAAATTTATGATCTGGCTCTGTCAGAAATTAATTTGGCGTAATGAATTAAGCAAGT 121

Sequence 8837, Application US/10767471
; GENERAL INFORMATION:

```

1  APPLICANT: CARGILL, Michele et al.
2  TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
3  TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
4  FILE REFERENCE: CLO04505
5  CURRENT APPLICATION NUMBER: US/10/767,471
6  CURRENT FILING DATE: 2004-01-30
7  NUMBER OF SEQ ID NOS: 50231
8  SOFTWARE: PasteSeq for Windows Version 4.0
9  SEQ ID NO: 8837
10 LENGTH: 201
11 TYPE: DNA
12 ORGANISM: Homo sapiens
13 us-10-767-471-8837

```

Query Match	6.8%;	Score 31.2;	DB 6;	Length 201;
Best Local Similarity	55.6%;	Pred. No. 1.8;		
Matches 60;	Conservative	0.0		

326 ACCGTAATAGTGGCA.com:www.poc... 0; Gaps 0
0; Mismatches 48; Indels

183 ATCCGCTTCTGGAATTAAGCCATATTACCCAGT 385

[illegible]

TCTGCGTCAGGACAGAAAGCTGG 76
RESULT 11

3-10-767-471-34439/c
Sequence 34439, Application US/10767471
GENERAL INFORMATION.

APPLICANT: CARGILL, Michele et al
TITLE OF INVENTION: GENETIC POLYM
TITLE OF INVENTION:

RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
 INVENTOR: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
 FILE REFERENCE: CL001505
 CURRENT APPLICATION NUMBER: US/10/767 473

CURRENT FILING DATE: 2004-01-30
NUMBER OF SEQ ID NOS: 50231
SOFTWARE: FastSeq for Windows Version 4.1

```

      SEQ ID NO 34439
      LENGTH: 201
      TYPE: DNA
      WINDOWS VERSION 4.0

```

ORGANISM: Homo sapiens
-10-767-471-34439

Query Match	6.8%	Score 31.2;	DB 6;	Length 201
Best Local Similarity	55.6%	Pred. No. 1.8;		
Matches 60;	Conservative			

[illegible]

183 ATGGAGTATGAGCGCAGATCCAGATGCTGATGATATATGAGAGCGCTTTTGTGTTT
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

386 ACCGTACGAGAAACAGCAGTTGTAGTCAGAAACAGTCGATATGG 433

123 AGGATTATCATGTACATGATGASGCAGAGCAGGAACAGAAAAAGCTGS 76

JLT 12
50-548-091-5725
Sequence 5725, Application US/60548091

```

1 GENERAL INFORMATION:
2 APPLICANT: CARGILL, Michele et al.
3 TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
4 TITLE OF INVENTION: STROKE, METHODS OF DETECTION AND USES THEREOF
5 FILE REFERENCE: CL001506

```

;; CURRENT APPLICATION NUMBER: US/60/548,091
;; CURRENT FILING DATE: 2004-02-27
;; NUMBER OF SEQ ID NOS: 2433

```

; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO 5725
;
; LENGTH: 394459

```

TYPE: DNA
ORGANISM: Homo sapiens

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(394468)

```

OTHER INFORMATION: n = A
US-60-548-091-5725

Query Match	6.8%	Score 31,	DB 7,	Length 394468,
Best Local Similarity	57.9%	Pred. No. 18:		
Matches				

[illegible][illegible]

293 ATGCCAAGATGCCAGTATATCGAAGCCCTTAC 327
 |||||
 320949 GTGTCAATAACCCAGTAGTCCCAAGGTTTCC 320983
 |||||

RESULT 13

Sequence 43, Application US/10257047
GENERAL INFORMATION:

APPLICANT: STEWARD, MICHAEL
APPLICANT: COX, VIVienne FRANCES
TITLE OF INVENTION: DNA IMMUNIZATION

FILE REFERENCE: 37945-0041
CURRENT APPLICATION NUMBER: US/10/257,047
CURRENT FILING DATE: 2010-03-01

PRIOR FILING DATE: 2002-10-08
PRIOR APPLICATION NUMBER: PCT/GB01/01599
PRIOR FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: PCT/GB01/01599

PRIOR APPLICATION NUMBER: GB 0008582.9
 PRIOR FILING DATE: 2000-04-08
 NUMBER OF SEQ ID NOS: 66

SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 43
LENGTH: 309

TYPE: DNA
ORGANISM: Plasmodium falciparum
10-257-047-43

Query Match	6.8%;	Score 30.8;	DB 6;	Length 309;
Best Local Similarity	53.3%;	Pred. No 27;		

65; Conservative	0; Mismatches	57; Indels	0; Gaps
------------------	---------------	------------	---------

2 TGAATAACAATTGTTATTTATGATGTGCACATACTGGGTGCCCTGGGATTGCAACCG 61
|||||
132 TGAGAACCTTAACCCTACCTTCCCTCCTGCTTCCTAAGTAACTGGAGTGGCCTGCTTCTC

CGACAAATTATGACTGGCTGGTTCAGATAATAATTTGGCGTAATAATGAATTTTCCCTCCTA
62 CGACAAATTATGACTGGCTGGTTCAGATAATAATTTGGCGTAATAATGAATTTTCCCTCCTA

192 CACCGAAGAGACTGTGTTCTAACGGAAGAAGACTTACTGCAATGTACTAAGCCAA 25

122	CT	12.
252	CT	25.

UT 14

0-257-047-46

Sequence 46, Application: US/10257047
GENERAL INFORMATION:
APPLICANT: STEWARD, MICHAEL
APPLICANT: COX, VIVIERNE FRANCES
TITLE OF INVENTION: DNA IMMUNIZATION VECTORS
FILE REFERENCE: 37945-.0041
CURRENT APPLICATION NUMBER: US/10/257,047
PRIOR FILING DATE: 2002-10-08
PRIOR APPLICATION NUMBER: PCT/GB01/01599
PRIOR FILING DATE: 2001-04-09
PRIOR APPLICATION NUMBER: GB 0006582.9
PRIOR FILING DATE: 2000-04-08
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 46
LENGTH: 309
TYPE: DNA
ORGANISM: Plasmodium falciparum
US-10-257-047-46

Query Match	6.8%;	Score 30.8;	DB 6;	Length 309;
Best Local Similarity	53.3%;	Pred. No. 2.7;		
Matches	65;	Conservative	0;	Mismatches 57;
			Indels	0;
			Gaps	0

Oy 2 TGAATAACAATTTGATTATGATGTGCACATACTCGGGTGGCCTCGGAGTTCACACCG 61
 ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Dd 132 TGAGAACCACCAACCTCCTAAGCTTAACGAAACACGCTGATCGACGCTTGACGCTAACGTG 192
 ||| ||| | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

0y 62 CGACAATTATGATCTGGCTGTTCAGATATTAATTTGGCGTAATGAATTAAGCAAGT 121

Db 192 CACCGAAGAGACTCTGGTTCTAACGGAAGAAGATTACTTCGAATGTACTAAGCCAGA 251

Qy	122	CT	123
Db	252	CT	253

RESULT 15

US-10-257-047-44
; Sequence 44, Application US/10257047

```

; APPLICANT: STEWARD, MICHAEL
; APPLICANT: COX, VIVIANNE FRANCES
; TITLE OF INVENTION: DNA IMMUNIZATION VECTORS

```

; CURRENT APPLICATION NUMBER: US/10/257,047
 ; CURRENT FILING DATE: 2002-10-08
 ; PRIOR APPLICATION NUMBER: PCT/GB01/01599
 ;

;; PRIOR APPLICATION NUMBER: GB 0008582.5
;; PRIOR FILING DATE: 2000-04-08
; NUMBER OF SEQ ID NOS: 66

```

; SEQ ID NO 44
; LENGTH: 3147
; TYPE: DNA

```

```

; FEATURE:
; OTHER INFORMATION: Descriptive
; OTHER INFORMATION: pvk104-01

```

Query Match	6.8%	Score 30.8;	DB 6;	Length 3147;
Best Local Similarity	53.3%	Pred. No. 5.3;		
Matches 65;	Conservative 0;	Mismatches 57;	Indels 0;	Gaps 0;

Qy 2 TGAACCAAACTGTTATTATGATGTGACAAATCTGGGTGCGCTGGGATTGCAACG 61
Db 198 TGAGAACCAAACTGTTACTGTACGAGAACCAACGGTGGATTGCGACGCTGACGCTAAGTG 25

Qy 62 CGACAAATTAATGATCTGGCTCGTTGCAATTAATTTGGCGTAATGAATTAAGCAAGT 121
Db 258 CACCGAAGAAGACTCTGGTTCTAACGGAAGAAGATTACTTCGAATGTACTTAAGCCAGA 317

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-3
Perfect score: 456
Sequence: 1 atgaactttaaaagtaga.....cgacgcctcacagctactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_NA_Main:*

- 1: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
- 2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
- 3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
- 4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
- 5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
- 6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
- 7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
- 8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
- 9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
- 10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
- 11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
- 12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
- 13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
- 14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
- 15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
- 16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
- 17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
- 18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.*
- 19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
- 20: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
- 21: /cgn2_6/ptodata/2/pna/US095_COMB.seq.*
- 22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
- 23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
- 24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
- 25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
- 26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
- 27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
- 28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
- 29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
- 30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
- 31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
- 32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
- 33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
- 34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
- 35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
- 36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.*
- 37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
- 38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
- 39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
- 40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq.*
- 41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
- 42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*
- 43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*

- 44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
- 45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
- 46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
- 47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
- 48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*
- 49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.*
- 50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.*
- 51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.*
- 52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.*
- 53: /cgn2_6/ptodata/2/pna/US106_COMB.seq.*
- 54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq.*
- 55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq.*
- 56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
- 57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
- 58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
- 59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
- 60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
- 61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
- 62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
- 63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
- 64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
- 65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
- 66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
- 67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
- 68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
- 69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
- 70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
- 71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
- 72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
- 73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
- 74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
- 75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
- 76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
- 77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
- 78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
- 79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.*
- 80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.*
- 81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
- 82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
- 83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
- 84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
- 85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
- 86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
- 87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
- 88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
- 89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
- 90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
- 91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
- 92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
- 93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
- 94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
- 95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
- 96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
- 97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
- 98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
- 99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
- 100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
- 101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
- 102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
- 103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
- 104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*
- 105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq.*
- 106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq.*
- 107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq.*
- 108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq.*
- 109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq.*
- 110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	456	100.0	456	23	US-09-543-407-3	Sequence 3, Appl1
2	453	99.3	453	92	US-60-352-946-1	Sequence 1, Appl1
3	453	99.3	453	101	US-60-444-371-1	Sequence 1, Appl1
4	452.8	99.3	675	13	US-08-978-978-3	Sequence 3, Appl1
5	452.8	99.3	675	31	US-09-741-873B-3	Sequence 3, Appl1
6	396	86.8	396	13	US-08-978-978-1	Sequence 1, Appl1
7	360	78.9	360	31	US-09-741-873B-1	Sequence 1, Appl1
8	276.8	60.7	477	17	US-09-252-691-172	Sequence 172, Appl
9	276.8	60.7	477	17	US-09-252-691C-172	Sequence 172, Appl
10	276.8	60.7	477	51	US-10-417-886-172	Sequence 172, Appl
11	254.4	55.8	456	6	US-08-233-642A-56	Sequence 56, Appl1
12	254.4	55.8	456	23	US-09-543-407-1	Sequence 1, Appl1
13	238.4	52.3	456	23	US-09-543-407-23	Sequence 23, Appl1
14	236.8	51.9	456	23	US-09-543-407-19	Sequence 19, Appl1
15	230.4	50.5	456	23	US-09-543-407-17	Sequence 17, Appl1
16	230.4	50.5	456	23	US-09-543-407-21	Sequence 21, Appl1
17	228.8	50.2	456	23	US-09-543-407-27	Sequence 27, Appl1
18	227.2	49.8	456	23	US-09-543-407-29	Sequence 29, Appl1
19	225.6	49.5	456	23	US-09-543-407-15	Sequence 15, Appl1
20	225.6	48.8	456	23	US-09-543-407-13	Sequence 13, Appl1
21	220.8	48.4	456	23	US-09-543-407-25	Sequence 25, Appl1
22	216	47.4	456	23	US-09-543-407-11	Sequence 11, Appl1
23	173.4	38.0	361	6	US-08-233-642A-54	Sequence 54, Appl1
24	100	21.9	100	45	US-10-146-492B-77	Sequence 77, Appl1
25	100	21.9	100	45	US-10-146-492B-78	Sequence 78, Appl1
26	100	21.9	100	45	US-10-146-492B-79	Sequence 79, Appl1
27	81.8	17.9	522	17	US-09-252-691-170	Sequence 170, Appl
28	81.8	17.9	522	17	US-09-252-691C-170	Sequence 170, Appl
29	81.8	17.9	522	51	US-10-417-886-170	Sequence 170, Appl
30	43.2	9.5	1083	1	PCT-US01-08631-12549	Sequence 12549, A
31	42.6	9.3	37551	81	US-60-245-901-6	Sequence 6, Appl1
32	42.6	9.3	37551	82	US-60-258-252-2	Sequence 2, Appl1
33	42.6	9.3	135859	1	PCT-US02-225766-8330	Sequence 8330, Ap
34	42.6	9.3	135859	47	US-60-423-586-24	Sequence 1855, Ap
35	42.6	9.3	135859	99	US-60-427-194-24	Sequence 24, Appl1
36	42.6	9.3	135259	99	US-60-427-194-24	Sequence 24, Appl1
37	41.6	9.1	561	17	US-09-252-691-171	Sequence 171, Appl
38	41.6	9.1	561	17	US-09-252-691C-171	Sequence 171, Appl
39	41.6	9.1	561	51	US-10-417-886-171	Sequence 171, Appl
40	39	8.6	1390	48	US-10-017-161-181	Sequence 1437, Ap
41	39	8.6	1390	48	US-10-292-799-1437	Sequence 1437, Ap
42	38.2	8.4	625	43	PCT-US01-08631-23586	Sequence 23586, A
43	38.2	8.4	32768	77	US-60-213-197-372	Sequence 24675, A
44	38.2	8.4	32768	77	US-60-213-197-372	Sequence 372, App
45	38.2	8.4	87681	82	US-60-258-275-339	Sequence 239, App

ALIGNMENTS

```

RESULT 1
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Dorian, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. Coli

```

US-09-543-407-3

```

Query Match      100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 8,9e-130;
Matches 456; Conservative 0; Mismatches 0;

```

```

Qy      1 ATGAACCTTTTAAAAAGTAGAAGCAATTGACGCAATGTAATTCTCCGGTAGCGCTCTGGCA 60
Db      1 ATGAACACTTTTAAAAAGTAGAAGCAATTGACGCAATGTAATTCTCCGGTAGCGCTCTGGCA 60
Qy      61 GGTGTTGTTCTCTCAGTAAGCGCGCGCGCGGTAAACAAGGTGTGTGGCGGTAAATATAGCGGC 120
Db      61 GGTGTTGTTCTCTCAGTAAGCGCGCGCGCGGTAAACAAGGTGTGTGGCGGTAAATATAGCGGC 120
Qy      121 CCAAAATTTGAGCTGTAACATTTTACCAGTACGGGTGGCGGTAACTCTGCACCTTGCTTGCAA 180
Db      121 CCAAAATTTGAGCTGTAACATTTTACCAGTACGGGTGGCGGTAACTCTGCACCTTGCTTGCAA 180
Qy      181 ACTGATGCCCGGTAACTCTGACTTGACTATTAACCAAGATGGCGCGGTAAATGTGTGCAGAT 240
Db      181 ACTGATGCCCGGTAACTCTGACTTGACTATTAACCAAGATGGCGCGGTAAATGTGTGCAGAT 240
Qy      241 GTTGTCAGAGCGCTCAGATGACAGCTCAATCGATCTGACCCAAAGTGGCTTCGGTAAACAC 300
Db      241 GTTGTCAGAGCGCTCAGATGACAGCTCAATCGATCTGACCCAAAGTGGCTTCGGTAAACAC 300
Qy      301 GCTACTCTTGATCACTGGAACGGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Db      301 GCTACTCTTGATCACTGGAACGGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Qy      361 GGCAACGGGTGCGCAGCTTGAACCAAGTCACTTACTCTCCGTCAACGGTACCTACAGTT 420
Db      361 GGCAACGGGTGCGCAGCTTGAACCAAGTCACTTACTCTCCGTCAACGGTACCTACAGTT 420
Qy      421 GGCCTTGGTAAACAAGCGACCGGCTCATCGTACTCTAA 456
Db      421 GGCCTTGGTAAACAAGCGACCGGCTCATCGTACTCTAA 456

RESULT 2
US-60-352-946-1      / Sequence 1, Application US/60352946
                        / GENERAL INFORMATION:
                        / APPLICANT: Hulgren, Scott
                        / APPLICANT: Malt, Chapman
                        / TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
                        / FILE REFERENCE: WSHU 2059
                        / CURRENT APPLICATION NUMBER: US/60/352,946
                        / NUMBER OF SEQ ID NOS: 10
                        / SOFTWARE: PatentIn version 3.0
                        / SEQ ID NO 1
                        / LENGTH: 453
                        / TYPE: DNA
                        / ORGANISM: Escherichia coli
US-60-352-946-1

```

Query Match	99.3%;	Score 453;	DB 92;	Length 453;
Best Local Similarity	100.0%;	Pred. No. 7.5e-129;		
Matches 453; Conservative	0;	Mismatches		

Qy	1	ATGAACCTTTTAAAGTAGAAGCAATTGCAGCAATGCTATTCTCCGGTAGCCCTCTGGCA	60
Db	1	ATGAACCTTTTAAAGTAGAAGCAATTGCAGCAATGCTATTCTCCGGTAGCCCTCTGGCA	60
Qy	61	GGTGTGTTCTCCAGTACGGCGCGCGGGTAAACAAGTGTGGCGGTAAATATAGCGGC	120
Db	61	GGTGTGTTCTCCAGTACGGCGCGCGGGTAAACAAGTGTGGCGGTAAATATAGCGGC	120
Qy	121	CCAATTCTGAGCTGAAACATTTTACAGTACGGTGGCGGTAACTCTGCACCTTGCTGCA	180
Db	121	CCAATTCTGAGCTGAAACATTTTACAGTACGGTGGCGGTAACTCTGCACCTTGCTGCA	180

QY 181 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 240
DB 181 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 240
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 300
DB 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 300
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 360
DB 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 360
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 420
DB 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 420
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453
DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

RESULT 3

US-60-444-371-1
; Sequence 1, Application US/60444371
; GENERAL INFORMATION:
; APPLICANT: Hultgren, Scott J
; TITLE OF INVENTION: BACTERIAL MODEL SYSTEM FOR AMYLOID FORMATION
; FILE REFERENCE: WSHU 2059.1
; CURRENT APPLICATION NUMBER: US/60/444,371
; CURRENT FILING DATE: 2003-01-31
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; TYPE: DNA
; ORGANISM: Escherichia coli
US-60-444-371-1

Query Match 99.3%; Score 453; DB 101; Length 453;
Best Local Similarity 100.0%; Pred. No. 7.5e-129;
Matches 453; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTCCTCCGGTAGCGCTCTGCA 60
DB 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTCCTCCGGTAGCGCTCTGCA 60
QY 61 GGTGTTGTTCTCAGTACGCGCGCGGGGTAAACACGCGTGGTGGCGTAAATATAGCGGC 120
DB 61 GGTGTTGTTCTCAGTACGCGCGCGGGGTAAACACGCGTGGTGGCGTAAATATAGCGGC 120
QY 121 CCAAAATTCAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTGCA 180
DB 121 CCAAAATTCAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTGCA 180
QY 181 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 240
DB 181 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 240
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 300
DB 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 300
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 360
DB 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 360
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 420
DB 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 420
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453
DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

DB 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 453

RESULT 4

US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8601723-1
; EARLIER FILING DATE: 1988-05-06
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-08-978-878-3

Query Match 99.3%; Score 452.8; DB 13; Length 675;
Best Local Similarity 99.6%; Pred. No. 9.9e-129;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTCCTCCGGTAGCGCTCTGCA 60
DB 83 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGATTCCTCCGGTAGCGCTCTGCA 142
QY 61 GGTGTTGTTCTCAGTACGCGCGCGGGGTAAACACGCGTGGTGGCGTAAATATAGCGGC 120
DB 143 GGTGTTGTTCTCAGTACGCGCGCGGGGTAAACACGCGTGGTGGCGTAAATATAGCGGC 202
QY 121 CCAAAATTCAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTGCA 180
DB 203 CCAAAATTCAGCTGAACATTTACAGTACGCTGCGGTAACTCTGCACTTGCTGCA 262
QY 181 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 240
DB 263 ACTGATGCCCGCTAAGTCTGATGATTAATACCGAGATGGCGGGGATATGATGAT 322
QY 241 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 300
DB 323 GTTGGTCAGGGCTCAGATGACAGCTCAATCGATCGACCCAACTGGCTTCGGTAAACAGC 382
QY 301 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 360
DB 383 GCTACTCTTGATCAGTGGAAACGGCAAAATTTGAAATGACGGTTAAACAGTTCCGTGCT 442
QY 361 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 420
DB 443 GGCAACGGGTGCTGACAGTTGACACGAGCTGATCTAATCTCCGTCGACGTCAGTCAAGTT 502
QY 421 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 456
DB 503 GGCTTTGGTAAACAAACGCGACCGCTCATGACATGAC 538

RESULT 5

US-09-741-873B-3
Sequence 3, Application US/09741873B
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873B
PRIOR APPLICATION NUMBER: 2003-04-04
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1992-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83) ..(538)
US-09-741-873B-3

Query Match
Best Local Similarity 99.3%; Score 452.8; DB 31; Length 675;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 ATGAACTTTAAAGTAGAGCAATTTGACAAATCTCTCGTAGCGCTGSCA 60
83 ATGAACTTTAAAGTAGAGCAATTTGACAAATCTCTCGTAGCGCTGSCA 142
61 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 120
143 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 202
121 CCAAAATTCGAGCTGAACATTTACAGTACGCGGTAACCTTGCACTTGCTGCA 180
203 CCAAAATTCGAGCTGAACATTTACAGTACGCGGTAACCTTGCACTTGCTGCA 262
181 ACTGATGCCCGTAACCTTGCACTTGCACTTGCACTTGCACTTGCACTTGCA 240
263 ACTGATGCCCGTAACCTTGCACTTGCACTTGCACTTGCACTTGCACTTGCA 322
241 GTTGTGAGGGCTCAGATGACGCTCAATGATCTGACCCAACTGCTCGTAACAGC 300
323 GTTGTGAGGGCTCAGATGACGCTCAATGATCTGACCCAACTGCTCGTAACAGC 382
301 GCTACTCTTGATCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 360
383 GCTACTCTTGATCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 442
361 GGCACCGGTGCTCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 420
443 GGCACCGGTGCTCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 502
421 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 456
503 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 538

RESULT 6

US-08-978-878-1
Sequence 1, Application US/08978878
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
FILE REFERENCE: 012889-081
CURRENT APPLICATION NUMBER: US/08/978,878
PRIOR APPLICATION NUMBER: 1997-11-26
PRIOR FILING DATE: 1988-05-06
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1989-05-06
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1992-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-10-05
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1995-06-28
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-08-978-878-1

Query Match
Best Local Similarity 86.8%; Score 396; DB 13; Length 396;
Matches 396; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

61 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 120
1 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 60
121 CCAAAATTCGAGCTGAACATTTACAGTACGCGGTAACCTTGCACTTGCTGCA 180
61 CCAAAATTCGAGCTGAACATTTACAGTACGCGGTAACCTTGCACTTGCTGCA 120
181 ACTGATGCCCGTAACCTTGCACTTGCACTTGCACTTGCACTTGCACTTGCA 240
121 ACTGATGCCCGTAACCTTGCACTTGCACTTGCACTTGCACTTGCACTTGCA 180
241 GTTGTGAGGGCTCAGATGACGCTCAATGATCTGACCCAACTGCTCGTAACAGC 300
301 GCTACTCTTGATCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 360
361 GCTACTCTTGATCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 300
241 GCTACTCTTGATCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 360
361 GGCACCGGTGCTCAGTGAACCGCAAAATTTGAAATGACGCTTAAACAGTTGGTGT 420
421 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 456
361 GGTGTTGTTCTCAGTACGCGCGCGGTAACACGCGTGGCGGTAATTAAGCGC 396

RESULT 7

US-09-741-873B-1
Sequence 1, Application US/09741873B
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation

FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873B
PRIOR FILING DATE: 2003-04-04
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1988-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 360
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873B-1

Query Match 79.9%; Score 360; DB 31; Length 360;
Best Local Similarity 100.0%; Pred. No. 4e-100;
Matches 360; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 61 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 120
DB 1 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 60
QY 121 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 180
DB 61 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 120
QY 181 ACTATGCCGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 240
DB 121 ACTATGCCGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 180
QY 241 GTTGGTCAGGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 300
DB 181 GTTGGTCAGGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 240
QY 301 GCTACTCTTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 360
DB 241 GCTACTCTTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 300
QY 361 GGCAACGGTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 420
DB 301 GGCAACGGTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 360

RESULT 8

US-09-252-691-172

Sequence 172; Application US/09252691B

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstock et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252,691B

CURRENT FILING DATE: 1999-02-18

NUMBER OF SEQ ID NOS: 11324

SEQ ID NO 172

LENGTH: 477

TYPE: DNA

ORGANISM: Enterobacter cloacae

US-09-252-691-172

Query Match 60.7%; Score 276.8; DB 17; Length 477;
Best Local Similarity 75.4%; Pred. No. 2.4e-74;

Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCCGTAAGCGCTTGCA 60
DB 22 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCCGTAAGCGCTTGCA 81
QY 61 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 120
DB 82 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 141
QY 121 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 180
DB 142 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 201
QY 181 ACTATGCCGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 240
DB 202 ACTATGCCGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 261
QY 241 GTTGGTCAGGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 300
DB 262 GTTGGTCAGGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 321
QY 301 GCTACTCTTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 360
DB 322 GCTACTCTTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 381
QY 361 GGCAACGGTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 420
DB 382 GGCAACGGTGTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 441
QY 421 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 456
DB 442 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 477

RESULT 9

US-09-252-691C-172

Sequence 172; Application US/09252691C

GENERAL INFORMATION:

APPLICANT: Keith G. Weinstock et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER

FILE REFERENCE: 107196.135

CURRENT APPLICATION NUMBER: US/09/252,691C

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/094,145

PRIOR FILING DATE: 1998-07-24

PRIOR APPLICATION NUMBER: US 60/074,787

PRIOR FILING DATE: 1998-02-18

NUMBER OF SEQ ID NOS: 11326

SEQ ID NO 172

LENGTH: 477

TYPE: DNA

ORGANISM: Enterobacter cloacae

US-09-252-691C-172

Query Match 60.7%; Score 276.8; DB 17; Length 477;
Best Local Similarity 75.4%; Pred. No. 2.4e-74;
Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCCGTAAGCGCTTGCA 60
DB 22 ATGAACCTTTTAAAGTAGAACAATTGACCAATGTAATCTCCGTAAGCGCTTGCA 81
QY 61 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 120
DB 82 GGTGTTGTTCTCTGACGAGCGCGGCGGTAAACAGCGGTGCGGTAAATATAGCGGC 141
QY 121 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 180
DB 142 CCAAAATCTGAGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 201
QY 181 ACTATGCCGCTGACGATTTACAGTACGCGGTGCGGTAACTGCACTTGCTGCA 240

```

Db      202 AGGAGCGCCGAGATTCTTAATTGACCATTTACCAGCATGTGGTGTAAGCGGCGCAAT 261
Qy      241 GTTGTGACGGGCTCAGATGACAGCTCAATCGATGACCCAAAGTGCTTGGTAACAGC 300
Db      262 GTTGGCGAGGGGCTGATGACAGTTCTATGATCTGGTCAAAAAGGCTTTGGTAACAGC 321
Qy      301 GCTACTCTGATGACGGAACGGCAAAATTTCTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Db      322 GGCACCATGACCAATGAAATGACAAAGCTCTGTTATCAAGGTAAAAAGATTCCGCGGC 381
Qy      361 GGCACCGGTGTGACGATTGACAGACGATGATCTAATCTCCGTCAGACGTTGACTGAGTT 420
Db      382 GGCACCGGGGGGGGGGTGACAGACAGACGATCCGGCTCAACGGTACTGATGACACAGGTT 441
Qy      421 GCGTTTGTGTAACAGCGGACCGCTCATCATGACTAA 456
Db      442 GCGTTTGGCAACAGCGGACCGCACACAGTACTGA 477

```

RESULT 10

```

US-10-417-886-172
; Sequence 172, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417.886
; CURRENT FILING DATE: 2003-04-17
; PRIOR APPLICATION NUMBER: US/09/252.691C
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/094,145
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; PRIOR FILING DATE: 1998-02-18
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-10-417-886-172

```

```

Query Match      60.7%; Score 276.8; DB 51; Length 477;
Best Local Similarity 75.4%; Pred. No. 2.4e-74;
Matches 344; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGTATTCTCCGTAGCGCTTGCA 60
Db      22 ATGAACCTTTAAAGTAAAGTAAAGCAATTCAGCAATCGTATTCTCCGTAGCGCTTGCA 81
Qy      61 GGTGTTGTTCTCTGATACGGGCGGGGTAAACACAGGTTGGTGGTAAATTAAGCGGC 120
Db      82 GGTGCGGTACCAATTTGGGCGGCGGCAAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 141
Qy      121 CCAAAATTTAGTGAACATTTACAGTACGTTGCGGTAACTCTGCACTTGGCTGCA 180
Db      142 CTTACTCAACCTGACCATTTACAGTACGTTGCGGCGGTAACTCTGCACTTGGCTGCA 201
Qy      181 ACTGATGCCGTAACTGATCTGATCTGATTAATCCAGATAGCGGCGGTAAATGTCAGAT 240
Db      202 AGGAGCGCCGAGATTCTTAATTGACCATTTACCAGCATGTGGTGTAAGCGGCGCAAT 261
Qy      241 GTTGTGACGGGCTCAGATGACAGCTCAATCGATGACCCAAAGTGCTTGGTAACAGC 300
Db      262 GTTGGCGAGGGGCTGATGACAGTTCTATGATCTGGTCAAAAAGGCTTTGGTAACAGC 321
Qy      301 GCTACTCTGATGACGGAACGGCAAAATTTCTGAAATGACGGTTAAACAGTTCCGTGTGT 360
Db      322 GGCACCATGACCAATGAAATGACAAAGCTCTGTTATCAAGGTAAAAAGATTCCGCGGC 381
Qy      361 GCGTTTGTGTAACAGCGGACCGCTCATCATGACTAA 456
Db      442 GCGTTTGGCAACAGCGGACCGCACACAGTACTGA 477

```

```

Db      382 GGCACCGGGGGGGGTGACAGACAGCGTCCGGTCAACGGTGACTGTGACACAGTT 441
Qy      421 GCGTTTGTAAACAACGGACCGGCTCATGACTGACTAA 456
Db      442 GCGTTTGGCAACAACGGACCGCACACAGTACTGA 477

```

RESULT 11

```

US-08-233-642A-56
; Sequence 56, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Clouthier, Sharon C.
; APPLICANT: Doran, James L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
; FILE OF INVENTION: BASED VACCINES
; NUMBER OF SEQUENCES: 58
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/233,642A
; FILING DATE: 26-APR-1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: King, Joshua
; REGISTRATION NUMBER: 35,570
; REFERENCE/DOCKET NUMBER: 920043.403C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; TELEEX: 3723836 SEEDANDBERRY
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 456 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..456
US-08-233-642A-56

```

```

Query Match      55.8%; Score 254.4; DB 6; Length 456;
Best Local Similarity 72.4%; Pred. No. 2e-67;
Matches 330; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

```

```

Qy      1 ATGAACCTTTAAAGTAAAGCAATTCAGCAATCGTATTCTCCGTAGCGCTTGCA 60
Db      1 ATGAACCTTTAAAGTAAAGTAAAGCAATTCAGCAATCGTATTCTCCGTAGCGCTTGCA 60
Qy      61 GGTGTTGTTCTCTGATACGGGCGGGGTAAACACAGGTTGGTGGTAAATTAAGCGGC 120
Db      61 GGTGTTGTTCTCTGATACGGGCGGGGTAAACACAGGTTGGTGGTAAATTAAGCGGC 120
Qy      121 CCAAAATTTAGTGAACATTTACAGTACGTTGCGGTAACTCTGCACTTGGCTGCA 180
Db      121 CCGAGCTCAAGTTAGCATTTATCAGTACGTTCCGCTTAACGCTGCCCTTGGCTGCA 180
Qy      181 ACTGATGCCGTAACTGATCTGATCTGATTAATCCAGATAGCGGCGGTAAATGTCAGAT 240
Db      181 AGCGATGCCGTAAATCTGAAACAGCATTAACCAAGCGGCTTATGTGTAACGGGCGCAT 240

```

QY 1 ATGAACCTTTTAAAGTGAACCAATGAGCAATCGTATTTCTCCGGTAGCGCTCGCA 60
Db 1 ATGAACCTTTTAAAGTGAACCAATGAGCAATCGTATTTCTCCGGTAGCGCTCGCT 60
QY 61 GGTGTGTTCTCAGTAGCGGCGGCGGTAACCAAGGTGTGGCGGTAAATAATAGCGG 120
Db 61 GCGCTGCTTCCCAATGGGCGGCGGTAATCATTAACGCGCGGCGCAATAGTCCGCG 120
QY 121 CCAAAATTCGAGCTGNAACATTTAACAGTACGGTGGCGGTAACTCTGCACTTGTGCA 180
Db 121 CCGAGCTCAACGTTAGCATTTATACGTACGGTCCGCTAACCGCTGCGCTTGTGTGCA 180
QY 181 ACTGATGCCGTAACCTCTGACTTGAATTAACCAAGCATGGCGGCGGTAAATGGTGCA 240
Db 181 AGCGATGCCCGTAATCTGAAGACATTAACCAAGCGGTATATGTAAACGCGCGGAT 240
QY 241 GTTGGTCAGGGCTCAGATACAGCTCAATCGATCTGAACCAAGTGGCTTCGTAACAGC 300
Db 241 GTAGGCCAGGGGCGGATATAGTACTATTTGAATCACTACAGATGGTTCACAAGAAATAT 300
QY 301 GCTACTCTTGATCAGTGAACGGCAAAATTTGGAATGACGGTTAAACAGTTCGGTGGT 360
Db 301 GGCACCATGACACAGTGAACCGTAAAACTCCGATATTAATCTGTGGCCAAATACGGGGT 360
QY 361 GGCACAGGTGCTGCAATTACCAAGACTGCATCTAATCTCCGTCAACGTGACTCAGGTT 420
Db 361 AATAACGCCCGCGCTGTTAATCAGACCGGATCTGATTTCCAGCGTAAATGGTGGCTCAGGTT 420
QY 421 GGCTTTGGTAACACGCGACCGCTCATCAGTAA 456

QY	1	ATGAAACTTTTAAAGTGAAGACCAATTTGAGCAATCGTATTCCTCCGGTAGCGCTTGCGCA	60
Db	1	ATGAAACTTTTAAAGTGAAGACCAATTTGAGCAATCGTATTCCTCCGGTAGCGCTTGCGCA	60
QY	61	GCGTGTGTTCTCTGACGAGCGCGCGCGGTAAACAACGCGTGTGCGCGGTAAATATATACGCGC	120
Db	61	GCGCGTGTTCACAAATGCGGCGCGCGCGGTATTCATTAACGCGCGCGCAATAGTTCGCGC	120
QY	121	CCAAATTTCTGAGCTGAACATTTTACCAGTACGGTGGCGGTACTCTTGCACTTGTCTTGCCA	180
Db	121	CCGAGCTCAACGTTAGAGATTTATAGTACGGTTCGCGTAAACGCTGCGCTGTGTCGCAA	180
QY	181	ACTGATGCCCGTAACTCTGACTTTCATTAACCCAGCATGCGCGCGGTATGGTGCAGAT	240
Db	181	AGCAGTCCCGCTTAATCTGAAACGACATTAACCCAGAGCGGTATTAAGTAAACGCGCGCAT	240
QY	241	GTTGGTCAAGGCGCTCGATGACAGCTCAATCATGATCTTGACCCCAACGTGCTTCGGTAAACGC	300
Db	241	GTAGGCGCAGGCGCGGATATATGATCTATTGTAACGTGACTCAAAATGGTTTCAGAAATTAAT	300
QY	301	GCTACTCTTGATCAATGGAACGGCAAAATTTCTGAATGACGGTTAAACAGTTCGGTGTGT	360
Db	301	GCACCAATCGACCAATGGAACGCTTAAAACTATATGATCAAGCTGTAAACCGGTGTGTAAAC	360
QY	361	GGAACCGGTGCTGACGTTGACCGAGACTGCATTAATCTCCGCTCAACGCTGACTCAGGTT	420
Db	361	CATGAATATGGACATGCAAAATCAGACCGGATTTGATTCAGAGGTAAATGTTGGGTCAAGGTT	420
QY	421	GCGTTTGGTAAACAGCGACCGGCTATCACTAGTACTAA	456
Db	421	GGTTTGGCAACACGCCACGGCTAACCAAGTATTTAA	456

RESULT 14
US-09-543-407-19
Sequence 19, Application US/095343407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.

;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543,407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 19
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
;; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-19

Query Match 51.9%; Score 236.8; DB 23; Length 456;
Best Local Similarity 70.0%; Pred. No. 5.4e-62;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAAGCAATTCAGCAATCGTATTCTCCGGTAGCGCTTGCA 60
Db 1 ATGAACCTTTTAAAGTAGAAGCAATTCAGCAATCGTATTCTCCGGTAGCGCTTGCA 60
QY 61 GGTGTTCTCTCAGTACGGGGGGGGGGTACCAAGCGTGGCGGTAATATAGCGGC 120
Db 61 GGTGTTCTCTCAGTACGGGGGGGGGGTACCAAGCGTGGCGGTAATATAGCGGC 120
QY 121 CCAAACTTCTGAGTGAACATTTACAGTACGGTGGCGGTAATCTGCACTTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTTACAGTACGGTGGCGGTAATCTGCACTTGCTGCA 180
QY 181 ACTGATGCCGTAATCTGACTGATTAATCCAGATGCGGGGTAATGTCAGAT 240
Db 181 AGCGATGCCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GTTGTGAGGGCTGAGTACAGTCAATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAAGCGAGGGTGGGTAATATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCTACTCTGATCAGTGAAGCAAGCAAAATTCGAATGACGGTTAAACAGTTGGTGT 360
Db 301 GCAACCATGACAGTGAAGCAAGCAAAATTCGAATGATGATGATGATGATGATGAT 360
QY 361 GGCACGCGTGTGAGTGAACAGTCAATCTTCCGTCACAGTCACTGAGTT 420
Db 361 AATAACCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTGCGTCAGTT 420
QY 421 GGGTTGGTAACAAAGGACCGCTCATCAGACTAA 456
Db 421 GGTTTGGCAACAAAGGACCGCTCATCAGACTAA 456

RESULT 15
US-09-543-407-17

;; Sequence 17, Application US/09543407
;; GENERAL INFORMATION:
;; APPLICANT: White, Aaron P.
;; APPLICANT: Doran, James L.
;; APPLICANT: Collinson, S. Karen
;; APPLICANT: Kay, William W.
;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543,407
;; CURRENT FILING DATE: 2000-04-05
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 17
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence

;; FEATURE:
;; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
;; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-17

Query Match 50.5%; Score 230.4; DB 23; Length 456;
Best Local Similarity 69.1%; Pred. No. 5.2e-60;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTAGAAGCAATTCAGCAATCGTATTCTCCGGTAGCGCTTGCA 60
Db 1 ATGAACCTTTTAAAGTAGAAGCAATTCAGCAATCGTATTCTCCGGTAGCGCTTGCA 60
QY 61 GGTGTTCTCTCAGTACGGGGGGGGGGTACCAAGCGTGGCGGTAATATAGCGGC 120
Db 61 GGTGTTCTCTCAGTACGGGGGGGGGGTACCAAGCGTGGCGGTAATATAGCGGC 120
QY 121 CCAAACTTCTGAGTGAACATTTACAGTACGGTGGCGGTAATCTGCACTTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTTACAGTACGGTGGCGGTAATCTGCACTTGCTGCA 180
QY 181 ACTGATGCCGTAATCTGACTGATTAATCCAGATGCGGGGTAATGTCAGAT 240
Db 181 AGCGATGCCGTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GTTGTGAGGGCTGAGTACAGTCAATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAAGCGAGGGTGGGTAATATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCTACTCTGATCAGTGAAGCAAGCAAAATTCGAATGACGGTTAAACAGTTGGTGT 360
Db 301 GCAACCATGACAGTGAAGCAAGCAAAATTCGAATGATGATGATGATGATGATGAT 360
QY 361 GGCACGCGTGTGAGTGAACAGTCAATCTTCCGTCACAGTCACTGAGTT 420
Db 361 AATAACCGCGCTGTTAATCAGACCGCATCTGATTCAGCGTAATGTGCGTCAGTT 420
QY 421 GGGTTGGTAACAAAGGACCGCTCATCAGACTAA 456
Db 421 GGTTTGGCAACAAAGGACCGCTCATCAGACTAA 456

Search completed: March 18, 2004, 02:46:48
Job time: 2925.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-3

Perfect score: 456

Sequence: 1 atgaacttttaaaagtaga.....cgaccgctcactgactactaa 456

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Pending Patente NA New:*

1: /cgn2_6/ptodata/2/pna/BCT_NEM_COMB_seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEM_COMB_seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEM_COMB_seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEM_COMB_seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEM_COMB_seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEM_COMB_seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEM_COMB_seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	452.8	99.3	675	US-09-741-873C-3	Sequence 3, Appli
2	396	86.8	396	US-09-741-873C-1	Sequence 58, Appli
3	34.2	7.5	79977	US-10-765-799-58	Sequence 1, Appli
4	34.2	7.5	186449	US-10-021-698A-715	Sequence 715, App
5	33	7.2	439	US-10-021-698A-1006	Sequence 1006, App
6	32	7.0	1355	US-10-045-674A-593	Sequence 593, App
7	30.8	6.8	290892	US-10-767-471-10809	Sequence 10809, App
8	30.4	6.7	945	US-10-767-701-7789	Sequence 7789, App
9	30.2	6.6	63693	US-60-548-091-5635	Sequence 5635, App
10	30.2	6.6	290892	US-10-767-471-10809	Sequence 10809, App
11	30	6.6	372	US-10-021-698A-1217	Sequence 1217, App
12	30	6.6	992	US-10-021-698A-1019	Sequence 1019, App
13	30	6.6	1203	US-10-021-698A-796	Sequence 796, App
14	29.6	6.5	9278	PCT-US04-02242-43	Sequence 43, Appli
15	29.6	6.5	1203	PCT-US04-02242-43	Sequence 13342, A
16	29.2	6.4	306	US-10-100-683-13342	Sequence 13343, A
17	29.2	6.4	306	US-10-100-683-13343	Sequence 13343, A
18	29.2	6.4	851	US-10-767-795-1598	Sequence 1598, App
19	29.2	6.4	2776	US-10-100-683-3409	Sequence 3409, App
20	29.2	6.4	4660	US-10-788-792-80	Sequence 80, Appli
21	29	6.4	507	PCT-US04-05654-1128	Sequence 1128, App
22	29	6.4	551	US-10-767-701-681	Sequence 681, App
23	29	6.4	753	US-10-767-701-9504	Sequence 9504, App
24	28.8	6.3	1350	PCT-US04-05654-863	Sequence 863, App
25	28.8	6.3	6705	US-09-032-438C-5	Sequence 5, Appli
26	28.8	6.3	6705	US-10-336-215A-5	Sequence 5, Appli

27	28.8	6.3	6705	US-10-340-097B-5	Sequence 5, Appli
28	28.8	6.3	6819	US-09-032-438C-2	Sequence 2, Appli
29	28.8	6.3	6819	US-10-336-215A-2	Sequence 2, Appli
30	28.8	6.3	6819	US-10-336-215A-2	Sequence 2, Appli
31	28.8	6.3	6819	US-10-340-097B-2	Sequence 2, Appli
32	28.8	6.3	7318	US-10-783-528-12	Sequence 12, Appli
33	28.8	6.3	7488	US-09-032-438C-1	Sequence 1, Appli
34	28.8	6.3	7783	US-10-336-215A-1	Sequence 1, Appli
35	28.8	6.3	7783	US-10-336-215A-1	Sequence 1, Appli
36	28.8	6.3	7783	US-10-340-097B-1	Sequence 1, Appli
37	28.6	6.3	435	US-10-021-698A-1216	Sequence 1216, App
38	28.6	6.3	704	PCT-US04-05654-1022	Sequence 1022, App
39	28.6	6.3	860	PCT-US04-05654-1018	Sequence 1018, App
40	28.6	6.3	906	PCT-US04-05654-1020	Sequence 1020, App
41	28.6	6.3	74279	US-10-765-790-16	Sequence 16, Appli
42	28.6	6.2	524	US-10-767-701-22290	Sequence 22290, A
43	28.2	6.2	654	US-10-767-701-4153	Sequence 4153, App
44	28.2	6.2	1252	PCT-US04-05654-877	Sequence 877, App
45	28.2	6.2	1606	PCT-US04-05654-2525	Sequence 2525, App

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As well As Its Preparation
FILE REFERENCE: 012868-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83) ..(538)
US-09-741-873C-3
Query Match 99.3%; Score 452.8; DB 5; Length 675;
Best Local Similarity 99.6%; Pred. No. 2.3e-145;
Matches 454; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTAGAAGCAATTCGATTCCTCGGTAGCGCTTGCA 60
DB 83 ATGAACCTTTAAAGTAGAAGCAATTCGATTCCTCGGTAGCGCTTGCA 142
QY 61 GGTGTGTTCTCGATCGGCGCGGCGGTAAACGCGTGTGCGGTAAATATGCGGC 120
DB 143 GGTGTGTTCTCGATCGGCGCGGCGGTAAACGCGTGTGCGGTAAATATGCGGC 202
QY 121 CCAATTCGTAGCAATTAACAGTACGATCGGTGCGGTACTGTGCACTTGTGCA 180

Dp	203	CCAAATCTGAGCTGAACATTTCACAGTACGAGGGGGGCTAACTCTGCACCTTGCTCGAA	267
Qy	181	ACTATATGCCCTAACTCTGACTTGACTATTAACCAAGACATGGGGGGGTATATGTGCAGAT	240
Dp	263	ACTATATGCCCTAACTCTGACTTGACTATTAACCAAGACATGGGGGGGTATATGTGCAGAT	322
Qy	241	GTTGTGTCAAGGCTCAGATGACAGCTCAATCGATCTGACCCAAACGTGGTGTGGTAAACG	300
Dp	323	GTTGTGTCAAGGCTCAGATGACAGCTCAATCGATCTGACCCAAACGTGGTGTGGTAAACG	382
Qy	301	GCTACTCTGTATCACTGTGGAGACGGCAAAATTTCTGAAATGACGGTTAAACAGTTGAGTGT	360
Dp	383	GCTACTCTGTATCACTGTGGAGACGGCAAAATTTCTGAAATGACGGTTAAACAGTTGAGTGT	442
Qy	361	GGCAACGGTCTGCACTTGAACCAAGCTGATCTAACTCTTCGTCAACGTGACTCAAGTT	420
Dp	443	GGCAACGGTCTGCACTTGAACCAAGCTGATCTAACTCTTCGTCAACGTGACTCAAGTT	502
Qy	421	GGCTTTGTGAACAACGGGACCGGTGTATCGTAACTAA	456
Dp	503	GGCTTTGTGAACAACGGGACCGGTGTATCGTAACTAA	538

```

RESULT 2
US-09-741-873C-1
; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As well as Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741.873C
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

```

	Query Match	86.8%	Score 396	DB 5	Length 396
	Best Local Similarity	100.0%	Pred. No. 5.2e-126		
	Matches 396	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	61	GGTGTGTTCCCACTAGCGCGCGCGGTAA	CCACGGTGTGTGCGCGTAATATATAGCGCG	120	
Db	1	GGTGTGTTCCCACTAGCGCGCGCGGTAA	CCACGGTGTGTGCGCGTAATATATAGCGCG	60	
QY	121	CCAAATTTCTGAGCTGAACAATTTTACAGTA	CGGTGCGGTAACTGTGCACTTGCAATTTGCGAA	180	
Db	61	CCAAATTTCTGAGCTGAACAATTTTACAGTA	CGGTGCGGTAACTGTGCACTTGCAATTTGCGAA	120	
QY	181	ACTGATGCCCGTAATCTGTACTTGACTATTAA	CCCGAGATGGCGGCGGTAATATGGTGCBAAT	240	
Db	121	ACTGATGCCCGTAATCTGTACTTGACTATTAA	CCCGAGATGGCGGCGGTAATATGGTGCBAAT	180	
QY	241	GTGTGTCAGGGCTCAGATGACAGCTCATTCAT	CTGACCAACGTGGCTTTCGTATACGC	300	

Db	181	GTGGTAGAGGCTCAATATACAGCTCAATCGATTCGACCCCAAGCGTGGCTGGTAACAGC	240
QY	301	GCATCTCTTGATCATGTGGAAACGGCAAAATTTCTGAATATGACGTTTAAACAGTTCCGATG	360
Db	241	GCATCTCTTGATCATGTGGAAACGGCAAAATTTCTGAATATGACGTTTAAACAGTTCCGATG	300
QY	361	GGCAACGGGTCGACGTTGACCAAGCTGCATCTTAATCTCTCCGTCAACGTACTCAGGTT	420
Db	301	GGCAACGGGTCGACGTTGACCAAGCTGCATCTTAATCTCTCCGTCAACGTACTCAGGTT	360
QY	421	GGCTTTGTAAACAACGCAACGCGCTCATCAGTAATAA	456
Db	361	GGCTTTGTAAACAACGCAACGCGCTCATCAGTAATAA	396

```

RESULT 3
US-10-765-790-58/c
; Sequence 58, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Hatvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1.657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; CURRENT FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 79977
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-58

```

[illegible]

TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021,698A
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO 715
LENGTH: 186449
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (40991)
FEATURE:
NAME/KEY: modified_base
LOCATION: (41047)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (41057)..(41156)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (70192)..(70291)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90076)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90262)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90290)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90306)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90313)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90316)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (90344)..(90443)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (93719)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (111786)..(111885)
OTHER INFORMATION: a, t, c or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (129353)..(129452)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (144859)..(144958)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (155813)..(155912)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (166141)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (166388)..(166487)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (172617)..(172716)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (177984)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (178024)..(178123)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (180005)..(180104)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (183290)..(183389)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-715

Query Match 7.5%; Score 34.2; DB 6; Length 186449;
Best Local Similarity 43.9%; Pred. No. 0.88; Indels 0; Gaps 0;
Matches 147; Conservative 0; Mismatches 188;

QY 46 GGTAGCGCTCGAGGAGTGTCTCAGTACGCGCGGCTAACCGAGTGTGGC 105
DB 148605 GGTATGATGTGTGTGATGATGATGATGATGATGATGATGATGAT 148664
QY 106 GGTAAATATAGCGGCCCAATTCTGAGCTGAACATTTACAGTACGCTGCGTAACTCT 165
DB 148665 GATGATGATGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148724
QY 166 GCATTCGCTCGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 225
DB 148725 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148784
QY 226 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 285
DB 148785 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148844
QY 286 GGTTCGTTAAGACGCTACTCTTGAATGATGATGATGATGATGATGATGATGATGAT 345
DB 148845 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148904
QY 346 AAACAGTTGCTGTGCGCAACGATGATGATGATGATGATGATGATGATGATGAT 380
DB 148905 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 148939

RESULT 5
US-10-021-698A-1006/c

Query Match	7.2%	Score 33;	DB 6;	Length 439;
Best Local Similarity	47.8%	Pred. No. 0.13;		
Matches	96;	Conservative	0;	Minimum

	46	105	165	225	285	345	405	465	525	585	645	705	765	825	885	945	1005	1065	1125	1185	1245	1305	1365	1425	1485	1545	1605	1665	1725	1785	1845	1905	1965	2025	2085	2145	2205	2265	2325	2385	2445	2505	2565	2625	2685	2745	2805	2865	2925	2985	3045	3105	3165	3225	3285	3345	3405	3465	3525	3585	3645	3705	3765	3825	3885	3945	4005	4065	4125	4185	4245	4305	4365	4425	4485	4545	4605	4665	4725	4785	4845	4905	4965	5025	5085	5145	5205	5265	5325	5385	5445	5505	5565	5625	5685	5745	5805	5865	5925	5985	6045	6105	6165	6225	6285	6345	6405	6465	6525	6585	6645	6705	6765	6825	6885	6945	7005	7065	7125	7185	7245	7305	7365	7425	7485	7545	7605	7665	7725	7785	7845	7905	7965	8025	8085	8145	8205	8265	8325	8385	8445	8505	8565	8625	8685	8745	8805	8865	8925	8985	9045	9105	9165	9225	9285	9345	9405	9465	9525	9585	9645	9705	9765	9825	9885	9945	10005	10065	10125	10185	10245	10305	10365	10425	10485	10545	10605	10665	10725	10785	10845	10905	10965	11025	11085	11145	11205	11265	11325	11385	11445	11505	11565	11625	11685	11745	11805	11865	11925	11985	12045	12105	12165	12225	12285	12345	12405	12465	12525	12585	12645	12705	12765	12825	12885	12945	13005	13065	13125	13185	13245	13305	13365	13425	13485	13545	13605	13665	13725	13785	13845	13905	13965	14025	14085	14145	14205	14265	14325	14385	14445	14505	14565	14625	14685	14745	14805	14865	14925	14985	15045	15105	15165	15225	15285	15345	15405	15465	15525	15585	15645	15705	15765	15825	15885	15945	16005	16065	16125	16185	16245	16305	16365	16425	16485	16545	16605	16665	16725	16785	16845	16905	16965	17025	17085	17145	17205	17265	17325	17385	17445	17505	17565	17625	17685	17745	17805	17865	17925	17985	18045	18105	18165	18225	18285	18345	18405	18465	18525	18585	18645	18705	18765	18825	18885	18945	19005	19065	19125	19185	19245	19305	19365	19425	19485	19545	19605	19665	19725	19785	19845	19905	19965	20025	20085	20145	20205	20265	20325	20385	20445	20505	20565	20625	20685	20745	20805	20865	20925	20985	21045	21105	21165	21225	21285	21345	21405	21465	21525	21585	21645	21705	21765	21825	21885	21945	22005	22065	22125	22185	22245	22305	22365	22425	22485	22545	22605	22665	22725	22785	22845	22905	22965	23025	23085	23145	23205	23265	23325	23385	23445	23505	23565	23625	23685	23745	23805	23865	23925	23985	24045	24105	24165	24225	24285	24345	24405	24465	24525	24585	24645	24705	24765	24825	24885	24945	25005	25065	25125	25185	25245	25305	25365	25425	25485	25545	2
--	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	---

```

/ RESULT 6
/ US-10-045-674A-593
/ Sequence 593, Application us/10045674A
/ GENERAL INFORMATION:
/ APPLICANT: LADNER, ROBERT C.
/ APPLICANT: COHEN, EDWARD H.
/ APPLICANT: NASTRI, HORACIO G.
/ APPLICANT: ROOKEY, KRISTIN L.
/ APPLICANT: HOET, RENE
/ APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
/ TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
/ TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
/ TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
/ FILE REFERENCE: D1AX/002 C1P2
/ CURRENT APPLICATION NUMBER: US/10/045,674A
/ CURRENT FILING DATE: 2001-10-25
/ PRIOR APPLICATION NUMBER: 06/198,069
/ PRIOR FILING DATE: 2000-04-17
/ PRIOR APPLICATION NUMBER: 09/837,306
/ NUMBER OF SEQ ID NOS: 635
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 593
/ LENGTH: 1355
/ TYPE: DNA
/ ORGANISM: Artificial Sequence

```

```

; FEATURE: Description of Artificial Sequence: M13-III
; OTHER INFORMATION:
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
US-10-045-674A-593

```

Query Match	7.0%;	Score 32;	DB 6;	Length 1355;
Best Local Similarity	25.4%;	Pred. No. 0.49;		
Matches	59;	Conservative	64;	Mismatches 105

[illegible]

```

RESULT 7
US-10-767-471-10809/C
; Sequence 10809, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: C1001505
; CURRENT APPLICATION NUMBER: US/10767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 10809
; LENGTH: 290892
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(290892)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10809

```

Query Match	6.8%;	Score 30.8;	DB 6;	Length 290892;
Best Local Similarity	46.3%;	Pred. No. 16;		
Matches 101; Conservative	0.0;	Microbial		

[illegible]

RESULT 8

QY 122 CAATTCTGAGCTGAACATTACCACTACGTCGCGTAACTCTGCACTTGTCTGCAAA 181
DB 280 ATGATGCTATAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 221
QY 182 CTGATCCCGTAACCTGCTGACTATTAACCAAGATGCGCGGTAATGCTGACATG 241
DB 220 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 161
QY 242 TTGCTGAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 301
DB 160 ATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 101
QY 302 CTACTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
DB 100 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 41
QY 362 GCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 395
DB 40 ATGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 7

RESULT 12

PCT-US04-05654-1019
Sequence 1019, Application PC/TUS0405654
GENERAL INFORMATION:

APPLICANT: Sherman, Bradley K
APPLICANT: Riechmann, Jose Luis
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Heard, Jacqueline E
APPLICANT: Haake, Volker
APPLICANT: Creelman, Robert A
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddie, James
APPLICANT: Dubeil III, Arnold N
APPLICANT: Pineda, Omaira
APPLICANT: Repetti, Peter
APPLICANT: Century, Karen
APPLICANT: Guterson, Neal
APPLICANT: Yu, Guo-Liang
APPLICANT: Brown, Pierre E
APPLICANT: Kumamoto, Roderick W
APPLICANT: Pilgrim, Marsha L
FILE REFERENCE: MBI-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
PRIOR FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/675,852
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO 1019
LENGTH: 992
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
FEATURE:
NAME/KEY: misc feature
LOCATION: (978)..(984)
OTHER INFORMATION: n is a, c, g, or t
FEATURE:
NAME/KEY: misc feature
LOCATION: (987)..(987)
OTHER INFORMATION: n is a, c, g, or t
PCT-US04-05654-1019

Query Match 6.6%; Score 30; DB 1; Length 992;
Best Local Similarity 61.5%; Pred. No. 2.1;
Matches 48; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 42 CTCGGTAGCGCTCTGACAGATGTTGTTCTCAGTACGGCGCGGTAACCAAGCTGG 101
DB 324 CTATGCGCGGAGACAGAGAGGTGTGAGGTGATGATGATGATGATGATGATGATG 383
QY 102 TGGCGTAAATATAGCGG 119
DB 384 TGGCGGTGCTATGCTGG 401

RESULT 13

US-10-021-698A-796/c
Sequence 796, Application US/10021698A
GENERAL INFORMATION:

APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN BERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSEF
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
FILE REFERENCE: 2976-4044US1
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO 796
LENGTH: 9278
TYPE: DNA
ORGANISM: Homo sapiens
US-10-021-698A-796

Query Match 6.6%; Score 30; DB 6; Length 9278;
Best Local Similarity 46.3%; Pred. No. 5.9;
Matches 99; Conservative 0; Mismatches 115; Indels 0; Gaps 0;

QY 46 GGTAAGGCTCTGACAGATGTTGTTCTCAGTACGGCGCGGTAACCAAGCTGG 105
DB 359 GGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 106 GGTAATATAGCGCGCAATTTCTGAGCTGAACATTTACAGTACGGCGGTAACCT 165
DB 299 GATGCTGACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 240
QY 166 GCACTTGTCTGCAACATGATGATGATGATGATGATGATGATGATGATGATGATG 225
DB 239 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 180
QY 226 GGTAATGCTGACATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 259
DB 179 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 146

RESULT 14

PCT-US04-02242-43
Sequence 43, Application PC/TUS0402242
GENERAL INFORMATION:

APPLICANT: Diversa Corporation
APPLICANT: Barton, Nelson; Robertson, Dan; Elkins, James; Chang, Kristine
TITLE OF INVENTION: ENZYMES AND NUCLEIC ACIDS ENCODING THEM AND METHODS FOR MAKING ANI
FILE REFERENCE: 56446-20118_40
CURRENT APPLICATION NUMBER: PCT/US04/02242
CURRENT FILING DATE: 2004-01-26
PRIOR APPLICATION NUMBER: 60/442,794
PRIOR FILING DATE: 2003-01-24
NUMBER OF SEQ ID NOS: 224
SOFTWARE: PatentIn version 3.1

SEQ ID NO 43
LENGTH: 1203
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
PCT-US04-02242-43

Query Match 6.5%; Score 29.6; DB 1; Length 1203;
Best Local Similarity 56.0%; Pred. No. 3.1;
Matches 56; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 36 CGTATTCCTCCGGTAGCGCTCTGGCAGAGTGTGCTTCTGCTAGTAGCGGCGGCGGTACCA 95
DB 228 CCGTGAACACCGGTCGCCCTGCGCTGTGGGTTACCTTAAGATGCCGAGCGCAACACTTA 287
QY 96 CGGTGGTGGCGGTAAATATATATAGCGGCCCAATTCTGAGCTG 135
DB 288 CGGCGGCGGACGTAATATCTCGAAACACGATGCGCTG 327

RESULT 15

US-10-100-683-13342
Sequence 13342, Application US/10100683
GENERAL INFORMATION:
APPLICANT: Rosen, et al.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: P5900
CURRENT APPLICATION NUMBER: US/10/100,683
CURRENT FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: US 60/040,162
PRIOR FILING DATE: 1997-03-07
PRIOR APPLICATION NUMBER: US 60/043,576
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,601
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,845
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,580
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,599
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,314
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,632
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,892
PRIOR FILING DATE: 1997-08-22
Remaining Prior Application data removed - See File Wrapper or PALM.
NUMBER OF SEQ ID NOS: 13468
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13342
LENGTH: 306
TYPE: DNA
ORGANISM: Homo sapiens
US-10-100-683-13342

Query Match 6.4%; Score 29.2; DB 6; Length 306;
Best Local Similarity 52.5%; Pred. No. 2.2;
Matches 64; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 128 CTGAGCTGAACATTACACAGTAGGCTGGCGGTAACCTTGCACTTCTGCAACTGATG 187
DB 124 CTCAGTAGTTCCTCTCACCTCGGCGACAGTGTGCACTTGTACAAAGCTCTTG 183
QY 188 CCCGTAAGCTGACTTGAAGTATACCAAGATGGGCGGTAAAGTGCAAGATGTGATC 247
DB 184 TCTTTTGTTCGATTGAAGTCCGCAACATAGCGTTTATGCTGTGAAGCTGTGAATC 243
QY 248 AG 249
||

DB 244 AG 245

Search completed: March 17, 2004, 08:25:34
Job time : 71.6301 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaaaaacattcgtctatc.....ctcgctgacacacgctaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patente NA Main:

```
1: /cgn2_6/prodata/2/pna/US06_PCTUS_COMB.seq.*
2: /cgn2_6/prodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/prodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/prodata/2/pna/US080_COMB.seq.*
5: /cgn2_6/prodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/prodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/prodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/prodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/prodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/prodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/prodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/prodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/prodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/prodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/prodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/prodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/prodata/2/pna/US092B_COMB.seq.*
18: /cgn2_6/prodata/2/pna/US093B_COMB.seq.*
19: /cgn2_6/prodata/2/pna/US093B_COMB.seq.*
20: /cgn2_6/prodata/2/pna/US094_COMB.seq.*
21: /cgn2_6/prodata/2/pna/US095A_COMB.seq.*
22: /cgn2_6/prodata/2/pna/US095B_COMB.seq.*
23: /cgn2_6/prodata/2/pna/US095C_COMB.seq.*
24: /cgn2_6/prodata/2/pna/US095D_COMB.seq.*
25: /cgn2_6/prodata/2/pna/US095A_COMB.seq.*
26: /cgn2_6/prodata/2/pna/US095B_COMB.seq.*
27: /cgn2_6/prodata/2/pna/US095C_COMB.seq.*
28: /cgn2_6/prodata/2/pna/US095D_COMB.seq.*
29: /cgn2_6/prodata/2/pna/US096B_COMB.seq.*
30: /cgn2_6/prodata/2/pna/US097A_COMB.seq.*
31: /cgn2_6/prodata/2/pna/US097B_COMB.seq.*
32: /cgn2_6/prodata/2/pna/US097C_COMB.seq.*
33: /cgn2_6/prodata/2/pna/US098A_COMB.seq.*
34: /cgn2_6/prodata/2/pna/US098B_COMB.seq.*
35: /cgn2_6/prodata/2/pna/US098C_COMB.seq.*
36: /cgn2_6/prodata/2/pna/US098D_COMB.seq.*
37: /cgn2_6/prodata/2/pna/US099A_COMB.seq.*
38: /cgn2_6/prodata/2/pna/US099B_COMB.seq.*
39: /cgn2_6/prodata/2/pna/US099C_COMB.seq.*
40: /cgn2_6/prodata/2/pna/US099D_COMB.seq.*
41: /cgn2_6/prodata/2/pna/US099E_COMB.seq.*
42: /cgn2_6/prodata/2/pna/US099F_COMB.seq.*
43: /cgn2_6/prodata/2/pna/US100A_COMB.seq.*
```

```
44: /cgn2_6/prodata/2/pna/US100B_COMB.seq.*
45: /cgn2_6/prodata/2/pna/US101A_COMB.seq.*
46: /cgn2_6/prodata/2/pna/US101B_COMB.seq.*
47: /cgn2_6/prodata/2/pna/US102A_COMB.seq.*
48: /cgn2_6/prodata/2/pna/US102B_COMB.seq.*
49: /cgn2_6/prodata/2/pna/US103A_COMB.seq.*
50: /cgn2_6/prodata/2/pna/US103B_COMB.seq.*
51: /cgn2_6/prodata/2/pna/US104A_COMB.seq.*
52: /cgn2_6/prodata/2/pna/US104B_COMB.seq.*
53: /cgn2_6/prodata/2/pna/US106_COMB.seq.*
54: /cgn2_6/prodata/2/pna/US107A_COMB.seq.*
55: /cgn2_6/prodata/2/pna/US107B_COMB.seq.*
56: /cgn2_6/prodata/2/pna/US6000_COMB.seq.*
57: /cgn2_6/prodata/2/pna/US6001_COMB.seq.*
58: /cgn2_6/prodata/2/pna/US6002_COMB.seq.*
59: /cgn2_6/prodata/2/pna/US6003_COMB.seq.*
60: /cgn2_6/prodata/2/pna/US6004_COMB.seq.*
61: /cgn2_6/prodata/2/pna/US6005_COMB.seq.*
62: /cgn2_6/prodata/2/pna/US6006_COMB.seq.*
63: /cgn2_6/prodata/2/pna/US6007_COMB.seq.*
64: /cgn2_6/prodata/2/pna/US6008_COMB.seq.*
65: /cgn2_6/prodata/2/pna/US6009_COMB.seq.*
66: /cgn2_6/prodata/2/pna/US6010_COMB.seq.*
67: /cgn2_6/prodata/2/pna/US6011_COMB.seq.*
68: /cgn2_6/prodata/2/pna/US6012_COMB.seq.*
69: /cgn2_6/prodata/2/pna/US6013_COMB.seq.*
70: /cgn2_6/prodata/2/pna/US6014_COMB.seq.*
71: /cgn2_6/prodata/2/pna/US6015_COMB.seq.*
72: /cgn2_6/prodata/2/pna/US6016_COMB.seq.*
73: /cgn2_6/prodata/2/pna/US6017_COMB.seq.*
74: /cgn2_6/prodata/2/pna/US6018_COMB.seq.*
75: /cgn2_6/prodata/2/pna/US6019_COMB.seq.*
76: /cgn2_6/prodata/2/pna/US6020_COMB.seq.*
77: /cgn2_6/prodata/2/pna/US6021_COMB.seq.*
78: /cgn2_6/prodata/2/pna/US6022_COMB.seq.*
79: /cgn2_6/prodata/2/pna/US6023A_COMB.seq.*
80: /cgn2_6/prodata/2/pna/US6023B_COMB.seq.*
81: /cgn2_6/prodata/2/pna/US6024_COMB.seq.*
82: /cgn2_6/prodata/2/pna/US6025_COMB.seq.*
83: /cgn2_6/prodata/2/pna/US6026_COMB.seq.*
84: /cgn2_6/prodata/2/pna/US6027_COMB.seq.*
85: /cgn2_6/prodata/2/pna/US6028_COMB.seq.*
86: /cgn2_6/prodata/2/pna/US6029_COMB.seq.*
87: /cgn2_6/prodata/2/pna/US6030_COMB.seq.*
88: /cgn2_6/prodata/2/pna/US6031_COMB.seq.*
89: /cgn2_6/prodata/2/pna/US6032_COMB.seq.*
90: /cgn2_6/prodata/2/pna/US6033_COMB.seq.*
91: /cgn2_6/prodata/2/pna/US6034_COMB.seq.*
92: /cgn2_6/prodata/2/pna/US6035_COMB.seq.*
93: /cgn2_6/prodata/2/pna/US6036_COMB.seq.*
94: /cgn2_6/prodata/2/pna/US6037_COMB.seq.*
95: /cgn2_6/prodata/2/pna/US6038_COMB.seq.*
96: /cgn2_6/prodata/2/pna/US6039_COMB.seq.*
97: /cgn2_6/prodata/2/pna/US6040_COMB.seq.*
98: /cgn2_6/prodata/2/pna/US6041_COMB.seq.*
99: /cgn2_6/prodata/2/pna/US6042_COMB.seq.*
100: /cgn2_6/prodata/2/pna/US6043_COMB.seq.*
101: /cgn2_6/prodata/2/pna/US6044_COMB.seq.*
102: /cgn2_6/prodata/2/pna/US6045_COMB.seq.*
103: /cgn2_6/prodata/2/pna/US6046_COMB.seq.*
104: /cgn2_6/prodata/2/pna/US6047_COMB.seq.*
105: /cgn2_6/prodata/2/pna/US6048_COMB.seq.*
106: /cgn2_6/prodata/2/pna/US6049_COMB.seq.*
107: /cgn2_6/prodata/2/pna/US6050_COMB.seq.*
108: /cgn2_6/prodata/2/pna/US6051_COMB.seq.*
109: /cgn2_6/prodata/2/pna/US6052_COMB.seq.*
110: /cgn2_6/prodata/2/pna/US6053_COMB.seq.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	456	100.0	456	23	US-09-543-407-4
2	332.8	73.0	456	23	US-09-543-407-2
3	290.2	63.6	561	17	US-09-252-691-171
4	290.2	63.6	561	17	US-09-252-691-171
5	290.2	63.6	561	17	US-09-252-691-171
6	100	21.9	100	45	US-10-146-492B-80
7	100	21.9	100	45	US-10-146-492B-80
8	100	21.9	100	45	US-10-146-492B-80
9	39.6	8.7	7218	8	US-08-466-194-14
10	38.2	8.4	391	18	US-09-332-782-15700
11	38.2	8.4	391	21	US-09-515-694-15700
12	37.8	8.4	4860	51	US-10-431-652-3221
13	37.8	8.3	2000	36	US-09-887-272A-5263
14	37.4	8.2	456	23	US-09-543-407-13
15	37.4	8.2	68667	39	US-09-949-016-17017
16	37.4	8.2	76668	103	US-06-466-412-82785
17	36.8	8.1	573	43	US-10-021-323-5919
18	36.8	8.1	573	82	US-06-255-619-3919
19	36.6	8.0	456	23	US-09-543-407-25
20	36.4	8.0	396071	103	US-06-465-241-51735
21	36.2	7.9	42918	103	US-06-466-412-85430
22	36.2	7.9	2105303	39	US-09-947-916-273
23	36	7.9	413	46	US-10-176-912-156
24	36	7.9	413	46	US-10-179-524-156
25	36	7.9	413	46	US-10-184-634-156
26	36	7.9	413	46	US-10-184-634-156
27	36	7.9	395143	103	US-06-466-412-87989
28	36	7.9	395143	103	US-06-466-412-87989
29	35.8	7.9	85690	29	US-09-692-412-72
30	35.8	7.9	85690	33	US-09-803-736-1052
31	35.8	7.9	85690	55	US-10-746-294-12
32	35.6	7.8	9493	70	US-06-144-351-3079
33	35.6	7.8	12178	26	US-09-620-392-28400
34	35.6	7.8	12178	30	US-09-702-134-10097
35	35.6	7.8	12178	33	US-09-815-264-80300
36	35.6	7.8	13036	26	US-09-620-392-47103
37	35.6	7.8	13036	30	US-09-702-134-22694
38	35.6	7.8	13036	33	US-09-815-264-64696
39	35.6	7.8	32948	26	US-09-620-392-40033
40	35.6	7.8	32948	30	US-09-702-134-20486
41	35.6	7.8	32948	33	US-09-815-264-66935
42	35.6	7.8	59719	39	US-09-949-016-15616
43	35.6	7.8	163183	107	US-60-500-337-18909
44	35.4	7.8	610	44	US-10-098-754-5821
45	35.4	7.8	837105	39	US-09-948-124-76

ALIGNMENTS

RESULT 1
US-09-543-407-4
Sequence 4, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 456
TYPE: DNA
ORGANISM: E. coli

US-09-543-407-4

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 99-125;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAAAACAATTTGTTATTTATGATGTTAACTAGTGGTGGCTGGGATTTGACGCC	60
DB	1	ATGAAAAACAATTTGTTATTTATGATGTTAACTAGTGGTGGCTGGGATTTGACGCC	60
QY	61	GCAGCAGTTATGATTTAGTAACTGATTAATTAATTTGCGGTAAATGATTAAG	120
DB	61	GCAGCAGTTATGATTTAGTAACTGATTAATTAATTTGCGGTAAATGATTAAG	120
QY	121	TCTTCAATTTATAGCAGCAGCATAATTTGTCAGCTGAGCTATATATGCTCAGTTA	180
DB	121	TCTTCAATTTATAGCAGCAGCATAATTTGTCAGCTGAGCTATATATGCTCAGTTA	180
QY	181	CGCAGGAGGCTCAAACTTTTGGCGTGTGGCCAGAAAGTAGTCAACCGGGCA	240
DB	181	CGCAGGAGGCTCAAACTTTTGGCGTGTGGCCAGAAAGTAGTCAACCGGGCA	240
QY	241	AAATTTGACCAACAGAGATTTAACTTGCATATATTGATCAGCGGAGTCCAA	300
DB	241	AAATTTGACCAACAGAGATTTAACTTGCATATATTGATCAGCGGAGTCCAA	300
QY	301	GATCCAGTATTTGCGAAGTCTTATGTTATGATGATGATTAATTCAGAAAGTTCT	360
DB	301	GATCCAGTATTTGCGAAGTCTTATGTTATGATGATGATTAATTCAGAAAGTTCT	360
QY	361	GGAATTAAGCAATTTTACACAGTATGTTACTCAAAAACGCAATTTGATGACAGA	420
DB	361	GGAATTAAGCAATTTTACACAGTATGTTACTCAAAAACGCAATTTGATGACAGA	420
QY	421	CAGTGCAGTAATGCTATTCGCGTACACAACTTAA	456
DB	421	CAGTGCAGTAATGCTATTCGCGTACACAACTTAA	456

RESULT 2
US-09-543-407-2
Sequence 2, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis
US-09-543-407-2

QY	1	ATGAAAAACAATTTGTTATTTATGATGTTAACTAGTGGTGGCTGGGATTTGACGCC	60
DB	1	ATGAAAAACAATTTGTTATTTATGATGTTAACTAGTGGTGGCTGGGATTTGACGCC	60
QY	61	GCAGCAGTTATGATTTAGTAACTGATTAATTAATTTGCGGTAAATGATTAAG	120
DB	61	GCAGCAGTTATGATTTAGTAACTGATTAATTAATTTGCGGTAAATGATTAAG	120
QY	121	TCTTCAATTTATAGCAGCAGCATAATTTGTCAGCTGAGCTATATATGCTCAGTTA	180

```

Db      121 TCTTCATTATACAGCGGCCATTATGTCAAGTCGACGAGTAATATAGTCCAGACTA 180
Qy      181 CGGAGGAGAGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      181 CGCAGGAAAGGATCAAACTATGTCCGTATTTTCAAGAAGAGGAAATATCGGGCG 240
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCCAC 300
Db      241 AAGTCGACGAGGAGGAAATTAATTAATCTTGATATATGAGCAACCGGCAATGCCAC 300
Qy      301 GATCCAGATATTTGCAAGGTGCTTATGTAATACTGCGATGATTAATCAGAAAGTTCT 360
Db      301 GATCCAGATATTTGCAAGGCTTAAAGGTATAGTGCAGTATTAATCAGAAAGTTCT 360
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAAGTCAAAAACGCGCAATTGATGACAGA 420
Db      361 GGAATTAAGCAATATTTACACAGTATGTAAGTCAAAAACGAGTGTGTAGTGCAGAAA 420
Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGGTTAA 456
Db      421 CAGTCGCAATGCTATTGCGGTGACACAGGTTAA 456

```

RESULT 3

```

US-09-252-691-171
; Sequence 171, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252.691B
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-171

```

```

Query Match      63.6%; Score 290.2; DB 17; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

Qy      1 ATGAAAAACAATGTTATTTATGATGTTAACTACTGGGTGCGCTGGGATTGCAACC 60
Db      106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGCGCTGGGTTGTATC 165
Qy      61 GCACAGGTTATGATTTAGCTAATTCAGATATACTCGGGTAAATGAATGATGATAG 120
Db      166 GCAGGTATTCAGATTTAGCCAGTTCTGAATATATTTTGGATTAATGAATTAAGTAA 225
Qy      121 TCTTCATTATACAGGACGACATATGTCAGCTGAGTGAATTAATGATGCTCAGTTA 180
Db      226 GCTTCATACATACAGGACGACATATGTCAGGCTTACAGAAATTAATTCGATGTA 285
Qy      181 CGGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      286 CGCAGAGACGGTTCTAAATGCTGTCCGTATTTCTCAGAGGGCGGGAATAACCGCGC 345
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 300
Db      346 AATGTTATCAGTCAGGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 405
Qy      301 GATCCAGATATTTGCAAGGTGCTTATGTAATACTGCGATGATTAATCAGAAAGTTCT 360
Db      406 GATCCAGATATTTAGCGAGGCTTTTGGCAACCGCGCATGATTAATCAGAAAGGCTCG 465
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAAGTCAAAAACGCGCAATTGATGACAGA 420
Db      466 GGTAAACGGGGAATTAATTAACAGTATGTAAGTCAAAAACGAGTGTGTAGTACAGAGA 525

```

```

Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGGTTA 455
Db      526 CAGTCGCAATGCTATTGCGGTATTCAAGTTA 560

```

RESULT 4

```

US-09-252-691C-171
; Sequence 171, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252.691C
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 171
; LENGTH: 561
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-171

```

```

Query Match      63.6%; Score 290.2; DB 17; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```

```

Qy      1 ATGAAAAACAATGTTATTTATGATGTTAACTACTGGGTGCGCTGGGATTGCAACC 60
Db      106 ATGAAAAACAACGTTGTTATGATGTTTACATTACTGGGTGCGCTGGGTTGTATC 165
Qy      61 GCACAGGTTATGATTTAGCTAATTCAGATATACTCGGGTAAATGAATGATGATAG 120
Db      166 GCAGGTATTCAGATTTAGCCAGTTCTGAATATATTTTGGATTAATGAATTAAGTAA 225
Qy      121 TCTTCATTATACAGGACGACATATGTCAGCTGAGTGAATTAATGATGCTCAGTTA 180
Db      226 GCTTCATACATACAGGACGACATATGTCAGGCTTACAGAAATTAATTCGATGTA 285
Qy      181 CGGAGGAGGCTCAAACTTTGGCGGTGTGGCGCAAGAGTAGTACACCGGGCA 240
Db      286 CGCAGAGACGGTTCTAAATGCTGTCCGTATTTCTCAGAGGGCGGGAATAACCGCGC 345
Qy      241 AAGATTGACGACGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 300
Db      346 AATGTTATCAGTCAGGAGAGATTAATTAACCTTGATATATGATGAGCGGGAGTGGCAAC 405
Qy      301 GATCCAGATATTTGCAAGGTGCTTATGTAATACTGCGATGATTAATCAGAAAGTTCT 360
Db      406 GATCCAGATATTTAGCGAGGCTTTTGGCAACCGCGCATGATTAATCAGAAAGGCTCG 465
Qy      361 GGTATTAAGCAATATTTACACAGTATGTAAGTCAAAAACGCGCAATTGATGACAGA 420
Db      466 GGTAAACGGGGAATTAATTAACAGTATGTAAGTCAAAAACGAGTGTGTAGTACAGAGA 525
Qy      421 CAGTCGCAATGCTATTGCGGTGACACAGGTTA 455
Db      526 CAGTCGCAATGCTATTGCGGTATTCAAGTTA 560

```

RESULT 5

```

US-10-417-886-171
; Sequence 171, Application US/10417886
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/10/417.886
; CURRENT FILING DATE: 2003-04-17

```

PRIOR APPLICATION NUMBER: US/09/252,691C
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/094,145
PRIOR FILING DATE: 1998-07-24
PRIOR APPLICATION NUMBER: US 60/074,787
PRIOR FILING DATE: 1998-02-18
NUMBER OF SEQ ID NOS: 11326
SEQ ID NO 171
LENGTH: 561
TYPE: DNA
ORGANISM: Enterobacter cloacae
US-10-417-886-171

Query Match 63.6%; Score 290.2; DB 51; Length 561;
Best Local Similarity 77.4%; Pred. No. 2.5e-75;
Matches 352; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 1 ATGAAAAA...ATTGTTATTTATGATGTTTAACTAGTGGTGGCTGGAGTTGACGCC 60
DB 106 ATGAAAAA...AGTTGTTATGATGTTTAACTAGTGGTGGCTGGAGTTGATC 165
QY 61 GCAGCAGTATGATTTAGCTATTTAGAAATTAATTTGCGGTAATGAATTGATGTAAG 120
DB 166 GCAGCAGTATGATTTAGCTATTTAGAAATTAATTTGCGGTAATGAATTGATGTAAG 225
QY 121 TCTTCATTTAATGAGCAGCAATTTGTCAGAGCTGGGCAATTAATGATGCTAGTTA 180
DB 226 GCTTCATACATCAGCAGCAATTTGTCAGAGCTGGGCAATTAATGATGCTAGTTA 285
QY 181 CGCAGCAGGAGCTCAAACTTTGGCGGTTGGCGAAGAGTAGTAGCAACCGGCA 240
DB 286 CGCAGCAGGAGCTCAAACTTTGGCGGTTGGCGAAGAGTAGTAGCAACCGGCA 345
QY 241 AAGATTGACAGCAGAGATATTAACCTGATATTTATGAGTGGCGGCAATGCGCAAC 300
DB 346 AATGTTGATCAGTACAGCAGATTAACCTGATATTTATGAGTGGCGGCAATGCGCAAC 405
QY 301 GATGCAATTTGCGAAGTGTCTTATGATATCTGATATTCAGAAAGTTCT 360
DB 406 GATGCAATTTGCGAAGTGTCTTATGATATCTGATATTCAGAAAGTTCT 465
QY 361 GGTATTAAGCAAAATTTATACAGATAGTACTCAAAAAGCGCAATTTAGTGCAGAGA 420
DB 466 GGTATTAAGCAAAATTTATACAGATAGTACTCAAAAAGCGCAATTTAGTGCAGAGA 525
QY 421 CAGTGCAGATGCTATTTGCGGTGACACAGCTTA 455
DB 526 CAGTGCAGATGCTATTTGCGGTGACACAGCTTA 560

RESULT 6
US-10-146-492B-80
Sequence 80, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 80
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 39-58
US-10-146-492B-80

Query Match 21.9%; Score 100; DB 45; Length 100;

Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 266 ACCTTGATATATGATCAGCGGCGGAGTGCACAGATGATTTGGCAAGTCTT 325
DB 1 ACCTTGATATATGATCAGCGGCGGAGTGCACAGATGATTTGGCAAGTCTT 60
QY 326 ATGGTAATACCTGCATATTTATCCAGAAAGTTCTGTTAA 365
DB 61 ATGGTAATACCTGCATATTTATCCAGAAAGTTCTGTTAA 100

RESULT 7
US-10-146-492B-81
Sequence 81, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 81
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 38-57
US-10-146-492B-81

Query Match 21.9%; Score 100; DB 45; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 208 GTTGTGGCGAAGAGTAGTAGCAACCGGCAAAAGATTGACAGAGGATTTATAC 267
DB 1 GTTGTGGCGAAGAGTAGTAGCAACCGGCAAAAGATTGACAGAGGATTTATAC 60
QY 268 CTTGATATTTGATCAGCGGCGGCAATGCGCAACATGCCA 307
DB 61 CTTGATATTTGATCAGCGGCGGCAATGCGCAACATGCCA 100

RESULT 8
US-10-146-492B-82
Sequence 82, Application US/10146492B
GENERAL INFORMATION:
APPLICANT: MMG Biotech AG
TITLE OF INVENTION: Biochip
FILE REFERENCE: REN-14043
CURRENT APPLICATION NUMBER: US/10/146,492B
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: EP 0111279.5
NUMBER OF SEQ ID NOS: 12816
SEQ ID NO 82
LENGTH: 100
TYPE: DNA
ORGANISM: Escherichia coli K-12 MG1655
FEATURE:
OTHER INFORMATION: csgb b1042 U00096 1103174..1103629
PUBLICATION INFORMATION:
RELEVANT RESIDUES: 57-76
US-10-146-492B-82

Query Match 21.9%; Score 100; DB 45; Length 100;
Best Local Similarity 100.0%; Pred. No. 6.7e-19;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 123 TTCAATTATGAGCAGCAGCATTAATGTCAGAGCTGAGTAAATAGTGCAGTTGCG 182
Db 1 TTCAATTATGAGCAGCAGCATTAATGTCAGAGCTGAGTAAATAGTGCAGTTGCG 60
Qy 183 GCAGGAGGCTCAAAACTTTTGGCGGTTGTTGGCGAAGAA 222
Db 61 GCAGGAGGCTCAAAACTTTTGGCGGTTGTTGGCGAAGAA 100

RESULT 9

US-08-466-194-14/c
; Sequence 14, Application US/08466194
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHERFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, NW, Suite 500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20007
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,194
; FILING DATE: 06-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,463
; FILING DATE: 22-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/935,313
; FILING DATE: 26-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/201 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZspc-F18
; US-08-466-194-14

Query Match 8.7%; Score 39.6; DB 8; Length 7218;
Best Local Similarity 10.5%; Pred. No. 2.8;
Matches 39; Conservative 166; Mismatches 165; Indels 0; Gaps 0;

Qy 64 GCAGCTATGATTTAGCTAATTCAGATATTAATTCGCGGTAATGAATTGAGTACT 123
Db 1523 GCATCTATTCAGTTTCAAAAAACGCGATGAGCATCTAATTAATCTATGCA 1464
Qy 124 TCATTTATGAGCAGCAGCATTAATGTCAGAGCTGAGTAAATAGTGCAGTTGCG 183
Db 1463 GTACGTTAAAGATGAAAGATTGTTGACRRRRRRRRRRRRRRRRRRRRRRRR 1404
Qy 184 CAGGAGGCTCAAAACTTTTGGCGGTTGTTGCGAAGAGTAGAGCAACGGGCAAG 243

Db 1403 RRR 1344
Qy 244 ATGACGACAGAGAGATTAATTAACCTTGACATATGTCAGCGGCGAGCCAGCAT 303
Db 1343 RRR 1284
Qy 304 GCCAGTATTTGCGAAGGTGCTTAATGTAATCTGCGATGATTAACGAAAGTTCTGT 363
Db 1283 RRR 1224
Qy 364 AATAAGCAATATTAACAGTATGTAATCAAAAAACGCAATTGTGACAGAGAGAG 423
Db 1223 RRR 1164
Qy 424 TCGCAATGCG 433
Db 1163 RRRRRRRRR 1154

RESULT 10

US-09-332-782-15700
; Sequence 15700, Application US/09332782
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-754CON1
; CURRENT APPLICATION NUMBER: US/09/332,782
; CURRENT FILING DATE: 1999-06-14
; EARLIER APPLICATION NUMBER: US 09/181,430
; EARLIER FILING DATE: 1998-10-28
; NUMBER OF SEQ ID NOS: 21027
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 15700
; LENGTH: 391
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-332-782-15700

Query Match 8.4%; Score 38.2; DB 18; Length 391;
Best Local Similarity 58.3%; Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

Qy 309 TATTCGCAAGGCTGTAATGTAATCTGCGATATTAATCCAGAAAGTTCTGTATTA 368
Db 277 TTTCGCCCAACAGAGCTGTGTCATGCTGAATGATATTCGCAATATCTGGAATGT 336
Qy 369 AGCAATATTAACAGATGTAATCTCAAAAAACGCAATTGTAGCAGAGACAG 423
Db 337 AGCCAGACTGCAATTAATGATGCAATTAATCAACACTTCTGTAGCATATATCG 391

RESULT 11

US-09-515-694-15700
; Sequence 15700, Application US/09515694
; GENERAL INFORMATION:
; APPLICANT: Arterburn, Matthew
; APPLICANT: Asghari, Vida
; APPLICANT: Blum, Linda
; APPLICANT: Cheung, Patrick
; APPLICANT: Danavandi, Simin
; APPLICANT: Dickson, Mark
; APPLICANT: Drake, Jim
; APPLICANT: Drmanac, Radoje
; APPLICANT: Engleman, Carrie
; APPLICANT: Faulkner, Brandy
; APPLICANT: Fox, Melvin
; APPLICANT: Garcia, Veronica
; APPLICANT: Gietek, Gretchen
; APPLICANT: Jomek, Leni
; APPLICANT: Jones, Lee
; APPLICANT: Kite, David

APPLICANT: Kofler, Janette
APPLICANT: Labat, Ivan
APPLICANT: Lee, Mon-Jae
APPLICANT: Lomelli, Michelle
APPLICANT: Nguyen, Hong
APPLICANT: Nguyen, Linh
APPLICANT: Nguyen, Lynne
APPLICANT: Nguyen, Phuong
APPLICANT: Ngira, Margie
APPLICANT: Ojeda, Jesse
APPLICANT: Palencia, Servando
APPLICANT: Ratsel, Fariba
APPLICANT: Randhwa, Gurpreet
APPLICANT: Sabourieh, Hannah
APPLICANT: Sidhu, Navjivan
APPLICANT: Smith, Benjamin
APPLICANT: Smythe, Ashleigh
APPLICANT: Tkach, Joe
APPLICANT: Tulpule, Mukul
APPLICANT: Verma, Ron
APPLICANT: Wachter, Adam
APPLICANT: Wu, James
APPLICANT: Yim, Kenneth
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FILE REFERENCE: 754C1P
CURRENT APPLICATION NUMBER: US/09/515,694
CURRENT FILING DATE: 2000-02-29
EARLIER APPLICATION NUMBER: 09/332,782
EARLIER FILING DATE: 1999-06-14
EARLIER APPLICATION NUMBER: 09/181,430
EARLIER FILING DATE: 1998-10-28
NUMBER OF SEQ ID NOS: 21027
SOFTWARE: fastseq for windows Version 3.0
SEQ ID NO 15700
LENGTH: 391
TYPE: DNA
ORGANISM: Homo sapiens
US-09-515-694-15700

Query Match
Best Local Similarity 8.4%; Score 38.2; DB 21; Length 391;
Best Local Similarity 58.3%; Pred. No. 2.7;
Matches 67; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 309 TATTTCGCAAGCTGTATGTAATCTGCGATGATATTCAGAAAGTTCTGTATTA 368
DB 277 TTTCGGCGCAAGCTGTGTGTCATCTGTAACGTGATATTCGCAATTAATCTGGAATGT 336

QY 369 AGCAATATATACAGATGATGTAATCAAAAACGCAATGATGAGAGAGAG 423
DB 337 AGCCAGACTGCAATATATATGACATATATCAGACACTTCTGTAGCATATCTG 391

RESULT 12
US-10-431-652-3221
Sequence 3221, Application US/10431652
GENERAL INFORMATION:
APPLICANT: Breton, Gary L.
APPLICANT: Bush, David
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: PATH03-08
CURRENT APPLICATION NUMBER: US/10/431,652
CURRENT FILING DATE: 2003-05-06
PRIOR APPLICATION NUMBER: US 09/328,352
PRIOR FILING DATE: 1999-06-04
PRIOR APPLICATION NUMBER: US 60/088,701
PRIOR FILING DATE: 1998-06-09
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 3221
LENGTH: 4860
TYPE: DNA
ORGANISM: Acinetobacter baumannii

US-10-431-652-3221

Query Match
Best Local Similarity 8.4%; Score 38.2; DB 51; Length 4860;
Best Local Similarity 45.9%; Pred. No. 6.3;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 106 AATGAATGATGATGCTTATTCATTTATTCAGGAGCCATTAATGTCAGCTGGACTAAT 165
DB 3445 AATGCTATTGTTAATTTATTTAGTACTGATGACGATTAATCTTACAGAGTGTCTGAG 3504

QY 166 AATAGTCTCAGTTACGAGGAGGCTCAAACTTTGGCGGTTTTCGCAAGAAGT 225
DB 3505 ATTTCACTATTATGTTGTTGGAGGCANATGATCATTCAGTGTGTGATGATGAT 3564

QY 226 AGTAGCAACCGGCAAGATTGACGACAGAGATTAATACCTGATATATGATCAG 285
DB 3565 TATCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3624

QY 286 GCGGAGAGTCCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 345
DB 3625 TATGTTGTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3684

QY 346 ATCCAGAAAGTTCTGTATTAATTAAGCAATATATACAGATATG 388
DB 3685 TTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3727

RESULT 13
US-09-887-272A-5263/C
Sequence 5263, Application US/09887272A
GENERAL INFORMATION:
APPLICANT: Hou, Yu-Ming
APPLICANT: Quan, Sheng
APPLICANT: Chang, Hui-Song
APPLICANT: Zhu, Tong
APPLICANT: Whitman, Steve
APPLICANT: Goff, Steve
APPLICANT: Glazebrook, Jane
APPLICANT: Chen, Wengqiong
APPLICANT: Katagiri, Fumitaki
APPLICANT: Xie, Zhiyi
APPLICANT: Tao, Yi
APPLICANT: Zou, Guangzhou
APPLICANT: Cooper, Bret
TITLE OF INVENTION: PLANT GENES INVOLVED IN DEFENSE AGAINST
FILE REFERENCE: PATH02
CURRENT APPLICATION NUMBER: US/09/887,272A
CURRENT FILING DATE: 2001-06-23
PRIOR APPLICATION NUMBER: 60/213,634
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/214,926
PRIOR FILING DATE: 2000-06-23
PRIOR APPLICATION NUMBER: 60/261,320
PRIOR FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 60/264,353
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 60/273,879
PRIOR FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: 09/887,271
PRIOR FILING DATE: 2001-06-22
NUMBER OF SEQ ID NOS: 6813
SOFTWARE: fastseq for windows Version 4.0
SEQ ID NO 5263
LENGTH: 2000
TYPE: DNA
ORGANISM: Oryza sativa
US-09-887-272A-5263

Query Match
Best Local Similarity 8.3%; Score 37.8; DB 36; Length 2000;
Best Local Similarity 13.1%; Pred. No. 6.2;
Matches 36; Conservative 121; Mismatches 118; Indels 0; Gaps 0;

```

? PRIOR APPLICATION NUMBER: 60/241,755
? PRIOR FILING DATE: 2000-10-20
? PRIOR APPLICATION NUMBER: 60/237,768
? PRIOR FILING DATE: 2000-10-03
? PRIOR APPLICATION NUMBER: 60/231,498
? PRIOR FILING DATE: 2000-09-08
? NUMBER OF SEQ ID NOS: 207012
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 17017
? LENGTH: 68667
? TYPE: DNA
? ORGANISM: Human
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: (1)...(68667)
? OTHER INFORMATION: n = A,T,C or G
US-09-949-016-17017

```

	Query Match	Best Local Similarity	8.2%;	Score 37.4;	DB 39;	Length 68667;
	Matches 65;	Conservative	0;	Mismatches 46;	Indels 0;	Gaps 0;
Oy	327	TGGAAATACGCGCATGTTATCCAGAAAGTCTGGTATTAAGCAAAATATTACACAGTA				386
Db	17497	TGCTCATCTGTGTATAATTTCTTAAAGTGAAGCAATTAATATAATAATAATTA				17438
Oy	387	TGGAATCTCAAAAAACGGCAATTGTAGTCAGAGACAGTCGCAATGCGTAT				437
Db	17437	TGATTATGAGATAGCAGCATGTGTGCTGTGAGATTAATTCCAAATTAATTT				17287

Qy	387	TCGTACTCAAAAACCGCAATTGTAGTCAGAGACAGTCGAAATGGCTAT	437
Db	17437	TCATATAGAAATAGCAGCATGTGCTGTATGATTTACTTCCAAATTAACITTT	17387

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds

(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-4

Perfect score: 456
Sequence: 1 atgaaacaaatcgtatc.....ctcgctgacacacgttaa 456Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA New: *
1: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.4	8.2	65558	US-10-765-790-103	Sequence 103, App
2	35	7.7	675	US-09-741-873C-3	Sequence 73, Appl
3	34.4	7.5	519599	US-10-765-790-73	Sequence 9911, App
4	34.2	7.5	887	US-10-767-701-9911	Sequence 139, App
5	33.8	7.4	1149	US-10-771-241-139	Sequence 10653, A
6	33.4	7.3	38719	US-10-765-790-43	Sequence 31198, A
7	32.2	7.1	261922	US-10-767-471-10653	Sequence 2325, App
8	31.6	6.9	512	US-10-767-701-31198	Sequence 36, Appl
9	31.4	6.9	525	US-10-417-884A-2325	Sequence 8836, App
10	31.4	6.9	1245	US-60-546-745-36	Sequence 34493, A
11	31.4	6.9	53332	US-10-767-471-8836	Sequence 501, App
12	30.8	6.8	201	US-10-767-471-34493	Sequence 10718, A
13	30.8	6.8	1227	US-10-767-471-501	Sequence 2544, App
14	30.8	6.8	1227	US-10-767-471-10718	Sequence 10805, A
15	30.8	6.8	76573	US-10-767-471-10718	Sequence 3515, App
16	30.4	6.7	1227	US-10-417-884A-2544	Sequence 7877, App
17	30.2	6.6	3529	US-10-779-543-9918	Sequence 8837, App
18	29.8	6.6	1790242	US-10-767-471-10805	Sequence 34439, App
19	29.8	6.5	600	US-60-545-213-3515	Sequence 10911, A
20	29.8	6.5	600	US-60-545-213-3515	Sequence 2111, App
21	29.6	6.5	201	US-10-767-471-8837	Sequence 10891, A
22	29.6	6.5	201	US-10-767-471-34439	
23	29.6	6.5	40000	US-10-767-471-10911	
24	29.6	6.5	122673	US-10-765-790-33	
25	29.4	6.4	2806	US-10-021-698A-2111	
26	29.4	6.4	74677	US-10-767-471-10891	

27	29.2	6.4	1080	US-10-451-467A-619	Sequence 619, App
28	29.2	6.4	134134	US-10-767-471-10743	Sequence 10743, App
29	29.2	6.4	189268	US-10-021-698A-697	Sequence 697, App
30	29.2	6.4	290892	US-10-767-471-10809	Sequence 18004, A
31	29	6.4	4447	US-10-767-701-18004	Sequence 17618, A
32	29	6.4	532	US-10-767-701-17618	Sequence 10858, A
33	29	6.4	13948	US-10-767-471-10858	Sequence 10705, A
34	29	6.4	16201	US-10-767-471-10705	Sequence 5603, App
35	29	6.4	45268	US-60-548-091-5603	Sequence 326, App
36	28.8	6.3	1829	US-10-775-169-326	Sequence 13618, A
37	28.8	6.3	1914	US-10-767-701-13618	Sequence 12012, A
38	28.8	6.3	20767	US-10-100-663-12012	Sequence 793, App
39	28.6	6.3	408	US-10-417-884A-793	Sequence 285, App
40	28.6	6.3	1056	US-10-781-014-285	Sequence 449, App
41	28.6	6.3	1056	US-10-781-014-449	Sequence 5675, App
42	28.6	6.3	78785	US-09-978-167A-3	Sequence 10675, A
43	28.6	6.3	100374	US-60-548-091-5675	Sequence 18788, A
44	28.6	6.3	189817	US-10-767-471-10675	
45	28.4	6.2	477	US-10-779-543-18788	

ALIGNMENTS

RESULT 1
US-10-765-790-103/C
Sequence 103, Application US/10765790
GENERAL INFORMATION:
APPLICANT: Bayer Healthcare LLC
APPLICANT: Beard, Chris
APPLICANT: Burgess, Chris
APPLICANT: Gannon, Allison
APPLICANT: Harvey, Jeanne
APPLICANT: Lechner, John F.
APPLICANT: Li, Zheng
TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
FILE REFERENCE: 1657/2035
CURRENT APPLICATION NUMBER: US/10/765,790
CURRENT FILING DATE: 2004-01-27
PRIOR APPLICATION NUMBER: US 10/737,082
PRIOR FILING DATE: 2003-12-16
NUMBER OF SEQ ID NOS: 300
SOFTWARE: PatentIn version 3.2
SEQ ID NO 103
LENGTH: 65558
TYPE: DNA
ORGANISM: Homo sapiens
US-10-765-790-103

Query Match 8.2%; Score 37.4; DB 6; Length 65558;
Best Local Similarity 58.6%; Pred. No. 0.2;
Matches 65; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

Qy 327 TGGTAATACGCGATGATTAATCCGAAAGGTTGTGTAATTAACCAATATTACACAGTA 386
Db 16503 TGCATCTCTGTGTAATATTTCTTAAGTGTAAGGAAATTAATTAATTAATTAATTA 16444
16443 TGAATGAGAAATAGACACAGTGTGCTGTAATGTTACTTCCAAATTAATCTT 16393

RESULT 2
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Ribonectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1

```

; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-09-741-873C-3
```

```

Query Match
Best Local Similarity 7.7%; Score 35; DB 5; Length 675;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```

Qy 416 AGAGACAGTCGCAATGCTATTCGCGTACACAA 450
Db 1 AGAGACAGTCGCAATGCTATTCGCGTACACAA 35
```

```

RESULT 3
; Sequence 73, Application US/10765790
; GENERAL INFORMATION:
; APPLICANT: Bayer Healthcare LLC
; APPLICANT: Beard, Chris
; APPLICANT: Burgess, Chris
; APPLICANT: Gannon, Allison
; APPLICANT: Harvey, Jeanne
; APPLICANT: Lechner, John F.
; APPLICANT: Li, Zheng
; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
; FILE REFERENCE: 1657/2035
; CURRENT APPLICATION NUMBER: US/10/765,790
; PRIOR FILING DATE: 2004-01-27
; PRIOR APPLICATION NUMBER: US 10/737,082
; PRIOR FILING DATE: 2003-12-16
; NUMBER OF SEQ ID NOS: 300
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73
; LENGTH: 519599
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-765-790-73
```

```

Query Match
Best Local Similarity 7.5%; Score 34.4; DB 6; Length 519599;
Matches 56; Conservative 0; Mismatches 36; Indels 0; Gaps 0;

Qy 263 ATAACTTGATATATGATCAGCGGCGGACAGTCACCAAGTATTTGGCAAGTG 322
Db 332273 ATAAATTTGATTTTGTGATCAGTACGATGGGCTATCTCTGACAGCATTTGGAAAAGG 332214
Qy 323 CTATAGTAACTGCGATGATTAATTCAGAAA 354
Db 332213 CTATAGTAACTGCGATGATTAATTCAGAAA 3322182
```

```

RESULT 4
US-10-767-701-9911
```

```

; Sequence 9911, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; PRIOR FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 9911
; LENGTH: 887
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS40598_1
US-10-767-701-9911
```

```

Query Match
Best Local Similarity 7.5%; Score 34.2; DB 6; Length 887;
Matches 72; Conservative 0; Mismatches 63; Indels 0; Gaps 0;
```

```

Qy 203 TGGCGGTTGTCGCAAGAGGTAGTACCAACCGGCAAGATTGACACAGAGATT 262
Db 433 TGTCAATGATTCCTTAAGGTGTGATACAAATTTGGGAAATCTGATTGTCCTCAGAG 492
```

```

Qy 263 ATAACTTGATATATGATCAGCGGCGGACAGTCACCAAGTATTTGGCAAGTG 322
Db 493 ATGGCTTAATGTAAGCTTAAGAACCAACCAAGTAATCCAGGTTCTTAAGATA 552
```

```

Qy 323 CTATAGTAACTG 337
Db 553 GCAATGGGGAATG 567
```

```

RESULT 5
; Sequence 139, Application US/10771241
; GENERAL INFORMATION:
; APPLICANT: Zyskind, Judith
; APPLICANT: Forsyth, R. Allyn
; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
; FILE REFERENCE: ELITRA.001C1
; CURRENT APPLICATION NUMBER: US/10/771,241
; PRIOR FILING DATE: 2004-02-03
; PRIOR APPLICATION NUMBER: 09/492,709
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: 60/117,405
; PRIOR FILING DATE: 1999-01-27
; NUMBER OF SEQ ID NOS: 485
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 139
; LENGTH: 1149
; TYPE: DNA
; ORGANISM: E. coli
US-10-771-241-139
```

```

Query Match
Best Local Similarity 7.4%; Score 33.8; DB 6; Length 1149;
Matches 62; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
```

```

Qy 259 GATTATACCTTGCAATATATGATCAGCGGCGGACAGTCACCAAGTATTTGGCAA 318
Db 442 GATATATACCTTAATATATGATCAGCGGCGGACAGTCACCAAGTATTTGGCAA 501
Qy 319 GGTCCTTATGATATCTGATGATTAATTCAGAAAAGTTCTGTATATA 367
Db 502 GGCACGAGTGTACAGTTCTTATATTAATTCAGAGCTTATGATATA 550
```

```

RESULT 6
```

FILING DATE: 17-Apr-2003
PRIOR APPLICATION DATA:

142 ATGGGAGTATGGAGCGGAGAAATCCAGATGGTGATGATAATATGAGGCGTATTCCAGG 833

Db 82 AGATTATCATGTACATGATGAGCAGACAGAAAGAAAGCTGG 35

RESULT 13

US-10-767-471-34493/c
; Sequence 34493, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34493
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-34493

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 201;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCATGATTATCCGAAAGTTCTGTGTAATTAAGCAATATTACACAGT 385

Db 142 ATGGAGATGAGCGCGGAATCCAAAGATGGATGATTAATRAAGAGGCTATTTCACAG 83

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCCCAATGG 433

Db 82 AGATTATCATGTACATGATGAGCAGACAGAAAGAAAGCTGG 35

RESULT 14

US-10-767-471-501/c
; Sequence 501, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 501
; LENGTH: 1227
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-767-471-501

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 1227;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCATGATTATCCGAAAGTTCTGTGTAATTAAGCAATATTACACAGT 385

Db 476 ATGGAGATGAGCGCGGAATCCAAAGATGGATGATTAATRAAGAGGCTATTTCACAG 417

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCCCAATGG 433

Db 416 AGATTATCATGTACATGATGAGCAGACAGAAAGAAAGCTGG 369

RESULT 15

US-10-767-471-10718/c
; Sequence 10718, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01505

; CURRENT APPLICATION NUMBER: US/10/767,471

; CURRENT FILING DATE: 2004-01-30

; NUMBER OF SEQ ID NOS: 50231

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 10718

; LENGTH: 76573

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (1)..(76573)

; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-

US-10-767-471-10718

Query Match

Best Local Similarity 54.6%; Score 30.8; DB 6; Length 76573;
Matches 59; Conservative 1; Mismatches 48; Indels 0; Gaps 0;

Qy 326 ATGGTAATACCTGCATGATTATCCGAAAGTTCTGTGTAATTAAGCAATATTACACAGT 385

Db 69821 ATGGAGATGAGCGCGGAATCCAAAGATGGATGATTAATRAAGAGGCTATTTCACAG 69762

Qy 386 ATGGTACTCAAAAAACGCAATTGTAGTCAGACAGACAGTCCCAATGG 433

Db 69761 AGATTATCATGTACATGATGAGCAGACAGAAAGAAAGCTGG 69714

Search completed: March 17, 2004, 08:25:37
Job time : 72.6301 secs.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 307.572 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-9
Perfect score: 48
Sequence: 1 tatgatcgcgtcgtaccgcg.....cccatgaaatggtcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Pending Patents NA Main:

```

1: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*

```

```

44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	48	100.0	48	23	US-09-543-407-9
2	48	100.0	78	23	US-09-543-407-52
3	48	100.0	78	23	US-09-543-407-53
4	48	100.0	456	23	US-09-543-407-11
5	48	100.0	456	23	US-09-543-407-13
6	48	100.0	456	23	US-09-543-407-15
7	48	100.0	456	23	US-09-543-407-17
8	48	100.0	456	23	US-09-543-407-19
9	48	100.0	456	23	US-09-543-407-21
10	48	100.0	456	23	US-09-543-407-23
11	48	100.0	456	23	US-09-543-407-25
12	48	100.0	456	23	US-09-543-407-27
13	48	100.0	456	23	US-09-543-407-29
14	40.6	84.6	78	23	US-09-543-407-48
15	40.6	84.6	78	23	US-09-543-407-49
16	29.4	61.2	3	23	US-07-965-673A-9
17	27.8	57.9	2547	48	US-10-282-122A-36154
18	26.8	55.8	107223	21	US-09-528-237A-850
19	24.6	51.3	1780	15	US-09-107-433-2456
20	24.6	51.3	2448	53	US-10-617-320-2456
21	24.6	51.3	2544	1	US-10-472-928-1519
22	24.6	51.3	2544	52	US-10-472-928-1519
23	24.6	51.3	2547	48	US-10-282-122A-37710
24	24.6	51.3	2547	53	US-10-640-833-77
25	24.6	51.3	2547	62	US-10-640-833-77
26	24.6	51.3	2547	62	US-10-640-833-77
27	24.6	51.3	2547	62	US-10-640-833-77
28	24.6	51.3	3657	46	US-60-029-960-47
29	24.6	51.3	3657	58	US-60-029-960-47
30	24.6	51.3	23131	62	US-60-061-998-613
31	24.6	51.3	216258	52	US-10-472-928-4979
32	24	50.0	409	24	US-09-565-240-34030
33	24	50.0	409	27	US-09-654-617-77662
34	24	50.0	409	29	US-09-684-016-77662
35	24	50.0	409	51	US-10-424-599-11975
36	24	50.0	2023	26	US-09-620-392-69478
37	24	50.0	5823	51	US-10-437-963-69258
38	24	50.0	12979	30	US-09-702-134-5750
39	24	50.0	12979	33	US-09-815-264-58710
40	24	50.0	13038	26	US-09-620-392-30380
41	23.6	49.2	390	21	US-09-521-640-44946
42	23.4	48.8	119	35	US-09-865-439A-100294
43	23.4	48.8	119	76	US-60-207-458-144215
44	23.2	48.3	310	27	US-09-654-617-237766
45	23.2	48.3	310	29	US-09-684-016-237766

ALIGNMENTS

RESULT 1
US-09-543-407-9
Sequence 9, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 48
TYPE: DNA
ORGANISM: Leishmania major

US-09-543-407-9

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 48;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 48
DB
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 48

RESULT 2

US-09-543-407-52/c
Sequence 52, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 52
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-52

QY
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 48
DB
48 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 1

RESULT 3

US-09-543-407-53
Sequence 53, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 53
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-53

QY
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 48
DB
1 TATGATCAGCTGTTACCCGTTGTTACCATGAATGGACATGCA 48

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 78;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 4

US-09-543-407-11

; Sequence 11, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-11

Query Match

100.0%; Score 48; DB 23; Length 456;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 48

Db 382 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 429

RESULT 5

US-09-543-407-13

; Sequence 13, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 13

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-13

Query Match

100.0%; Score 48; DB 23; Length 456;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 48

Db 352 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 399

RESULT 6

US-09-543-407-15

; Sequence 15, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 15

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-15

Query Match

100.0%; Score 48; DB 23; Length 456;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 48

Db 67 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 114

RESULT 7

US-09-543-407-17

; Sequence 17, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

; FILE REFERENCE: 920043.406

; CURRENT APPLICATION NUMBER: US/09/543,407

; CURRENT FILING DATE: 2000-04-05

; NUMBER OF SEQ ID NOS: 59

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 17

; LENGTH: 456

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

; OTHER INFORMATION: sequence containing the replacement fragment

; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-17

Query Match

100.0%; Score 48; DB 23; Length 456;

Best Local Similarity 100.0%; Pred. No. 1.7e-08;

Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy

1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 48

Db 127 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAATGGACATGCA 174

RESULT 8

US-09-543-407-19

; Sequence 19, Application US/09543407

; GENERAL INFORMATION:

; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.

; APPLICANT: Collinson, S. Karen

; APPLICANT: Kay, William W.

;; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
;; FILE REFERENCE: 920043.406
;; CURRENT APPLICATION NUMBER: US/09/543,407
;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 19
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-19

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 48
DB 196 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 243

RESULT 9
US-09-543-407-21
; Sequence 21, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 21
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 48
DB 262 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 309

RESULT 10
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05

;; NUMBER OF SEQ ID NOS: 59
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 23
;; LENGTH: 456
;; TYPE: DNA
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
;; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-23

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 48
DB 331 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 378

RESULT 11
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-25

Query Match
Best Local Similarity 100.0%; Score 48; DB 23; Length 456;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 48
DB 172 TATGATCAGCTGCTTACCCGCTGTTGTTACCCATGAATGGCACATGCA 219

RESULT 12
US-09-543-407-27
; Sequence 27, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 456
; TYPE: DNA

ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 48
Db 241 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 288

RESULT 13
US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

Query Match 100.0%; Score 48; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 48
Db 307 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGCA 354

RESULT 14
US-09-543-407-48/c
Sequence 48, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 48
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-48

Query Match 84.6%; Score 40.6; DB 23; Length 78;
Best Local Similarity 91.5%; Pred. No. 1.1e-05;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47
Db 48 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 2

RESULT 15
US-09-543-407-49
Sequence 49, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 49
LENGTH: 78
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: PCR primer
US-09-543-407-49

Query Match 84.6%; Score 40.6; DB 23; Length 78;
Best Local Similarity 91.5%; Pred. No. 1.1e-05;
Matches 43; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47
Db 1 TATGATCAGCTGCTTACCCGTTGTTTACCCATGAAATGGCAGATGC 47

Search completed: March 18, 2004, 02:46:57
Job time : 311.872 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 7.32948 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-9

Perfect score: 48
Sequence: 1 tatgatcgcgtgtaccgcg.....cccatgaatgcacatgca 48

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_MA_New.*
1: /cgn2_6/ptodata/2/pna/PCR_NEW_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	23.2	48.3	435	US-10-767-795-1394	Sequence 1394, Ap
2	23.2	48.3	539	US-10-767-795-1392	Sequence 1392, Ap
3	23.2	48.3	545	US-10-767-701-29396	Sequence 29396, A
4	23.2	48.3	743	US-10-767-701-11493	Sequence 11493, A
5	22.2	46.2	1035	US-10-781-014-1167	Sequence 1167, App
6	22	45.8	808	US-10-779-543-5864	Sequence 5864, A
7	21.6	45.0	201	US-10-767-471-34311	Sequence 34311, A
8	21.6	45.0	2895	US-10-767-701-14458	Sequence 14458, A
9	21.6	45.0	546025	US-10-767-471-10717	Sequence 10717, A
10	21.4	44.6	607	US-10-767-701-31393	Sequence 31393, A
11	21	43.8	26377	US-10-765-790-41	Sequence 41, App1
12	20.8	43.3	175737	US-10-783-271-34	Sequence 34, App1
13	20.6	42.9	421	US-10-779-543-15761	Sequence 15761, A
14	20.6	42.9	905	US-10-781-469-49	Sequence 49, App1
15	20.6	42.9	198161	US-10-775-169-52	Sequence 52, App1
16	20.4	42.5	201	US-10-767-471-11899	Sequence 11899, A
17	20.4	42.5	1404	US-10-417-884A-2521	Sequence 2521, Ap
18	20.4	42.5	318752	US-10-767-471-10553	Sequence 10553, A
19	20.2	42.1	532	US-10-767-701-30765	Sequence 30765, A
20	20.2	42.1	779	US-10-781-014-169	Sequence 169, App
21	20.2	42.1	1203	US-10-417-884A-2612	Sequence 2612, App
22	20.2	42.1	4503	US-10-417-375A-31	Sequence 31, App1
23	20.2	42.1	106776	US-10-417-375A-30	Sequence 30, App1
24	20	41.7	596	US-10-767-701-657	Sequence 657, App
25	20	41.7	746	US-10-779-543-5903	Sequence 5903, Ap
26	20	41.7	863	US-10-767-701-8642	Sequence 8642, Ap

27	20	41.7	3213	6	US-10-417-884A-62	Sequence 62, App1
28	20	41.7	36135	6	US-10-100-683-11827	Sequence 11827, A
29	19.8	41.2	390	6	US-10-779-543-12187	Sequence 12187, A
30	19.8	41.2	594	6	US-10-417-884A-380	Sequence 380, App
31	19.8	41.2	600	7	US-60-545-213-391	Sequence 391, App
32	19.8	41.2	600	7	US-60-545-213-392	Sequence 392, App
33	19.8	41.2	600	7	US-60-545-213-542	Sequence 542, App
34	19.8	41.2	600	7	US-60-545-213-4663	Sequence 4663, App
35	19.8	41.2	600	7	US-60-545-213-4664	Sequence 4664, App
36	19.8	41.2	600	7	US-60-545-213-4814	Sequence 4814, App
37	19.8	41.2	757	6	US-10-100-683-5410	Sequence 5410, App
38	19.8	41.2	844	6	US-10-767-701-10879	Sequence 10879, A
39	19.8	41.2	7652	1	PCT-US03-31941A-1	Sequence 1, App1
40	19.8	41.2	8684	1	PCT-US03-31941A-19	Sequence 2, App1
41	19.8	41.2	10011	1	PCT-US03-31941A-19	Sequence 19, App1
42	19.6	40.8	201	6	US-10-767-471-34488	Sequence 34488, A
43	19.6	40.8	201	6	US-10-767-471-34490	Sequence 34490, A
44	19.6	40.8	201	6	US-10-767-471-34556	Sequence 34556, A
45	19.6	40.8	1013	6	US-10-767-701-12512	Sequence 12512, A

ALIGNMENTS

RESULT 1
US-10-767-795-1394
; Sequence 1394, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53534)B
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1394
; LENGTH: 435
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-gstronu33B027f06b1
US-10-767-795-1394

Query Match 48.3%, Score 23.2; DB 6; Length 435;
Best Local Similarity 70.5%; Pred. No. 1.6;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 3 TGATCAGCTGGTTACCGGTGTTACCCAGGAATGCACATG 46
DB 73 TGATTAACCTATTAACCTGTTGTTACCGAAGAACACATG 116

RESULT 2
US-10-767-795-1392
; Sequence 1392, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Zhou, Yihua
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53534)B
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596
; SEQ ID NO 1392
; LENGTH: 539
; TYPE: DNA
; ORGANISM: Gossypium hirsutum
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-gstronu33B027f06b1

US-10-767-795-1392

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 539;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 3 TGATCAGCTGGTACCCGCTGTGTTACCCATGAATGACATG 46
Db 29 TGATAACTCATACCTCCGTTGGTTACCAAGAGACATG 72

RESULT 3

US-10-767-701-29396/c

Sequence 29396, Application US/10767701

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53535)B

CURRENT APPLICATION NUMBER: US/10/767,701

NUMBER OF SEQ ID NOS: 2004-01-29

SEQ ID NO 29396

LENGTH: 545

TYPE: DNA

ORGANISM: Sorghum bicolor

FEATURE:

OTHER INFORMATION: Clone ID: 9299890

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 545;
Matches 28; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 5 ATCAGCTGTTACCCGCTGTGTTACCCATGAATG 40
Db 532 ATCAGCTGTTACCCGCTGTGTTACCCATGAATG 497

RESULT 4

US-10-767-701-11493

Sequence 11493, Application US/10767701

GENERAL INFORMATION:

APPLICANT: Kovalic, David K.

APPLICANT: Zhou, Yihua

TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

FILE REFERENCE: 38-21(53535)B

CURRENT APPLICATION NUMBER: US/10/767,701

NUMBER OF SEQ ID NOS: 2004-01-29

SEQ ID NO 11493

LENGTH: 743

TYPE: DNA

ORGANISM: Sorghum bicolor

FEATURE:

OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS45512_1

Query Match

Best Local Similarity 48.3%; Score 23.2; DB 6; Length 743;
Matches 31; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTACCCGCTGTGTTACCCATGAATGACAT 45
Db 269 ATGATCAGCTGGTACCCGCTGTGTTACCCATGAATGACAT 312

RESULT 5

US-10-781-014-167

Sequence 167, Application US/10781014

GENERAL INFORMATION:

APPLICANT: Pompeius, Markus

APPLICANT: Kroeger, Burkhard

APPLICANT: Schroder, Hartwig

APPLICANT: Zelder, Oskar

APPLICANT: Habenhauer, Gregor

TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS

TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY

FILE REFERENCE: BGI-126CPN

CURRENT APPLICATION NUMBER: US/10/781,014

PRIOR APPLICATION NUMBER: 2004-02-17

PRIOR FILING DATE: 2000-06-23

PRIOR FILING DATE: 1999-06-25

PRIOR FILING DATE: 1999-07-09

PRIOR FILING DATE: 1999-08-31

PRIOR FILING DATE: 1999-08-31

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

PRIOR FILING DATE: 1999-07-08

```

; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 09/297,648
; PRIOR FILING DATE: 2000-04-10
; PRIOR APPLICATION NUMBER: PCT/US99/01619
; PRIOR FILING DATE: 1999-01-28
; PRIOR APPLICATION NUMBER: 60/072,910
; PRIOR FILING DATE: 1998-01-28
; PRIOR APPLICATION NUMBER: 60/075,954
; PRIOR FILING DATE: 1998-02-24
; PRIOR APPLICATION NUMBER: 60/080,114
; PRIOR FILING DATE: 1998-03-31
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 23767
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5864
; LENGTH: 808
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 3, 16, 20, 57, 68, 77, 78, 80, 89, 105, 107, 108, 115, 117,
; LOCATION: 120, 123, 129, 133, 138, 140, 143, 147, 148, 156, 159, 163,
; LOCATION: 166, 172, 173, 174, 180, 186, 192, 197, 198, 201, 207, 209,
; LOCATION: 214, 216, 217, 218, 219, 226, 228, 236, 237, 238, 243
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 244, 245, 246, 254, 283, 308, 311, 317, 325, 330, 332, 339,
; LOCATION: 357, 358, 375, 385, 398, 401, 415, 422, 424, 426, 430, 434,
; LOCATION: 442, 448, 449, 451, 470, 471, 475, 476, 479, 480, 491,
; LOCATION: 497, 507, 510, 527, 559, 574, 576, 584, 591, 609, 636
; OTHER INFORMATION: n = A,T,C or G
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 639, 641, 643, 650, 660, 672, 673, 681, 684, 722, 743, 754,
; LOCATION: 762, 783, 784, 795, 801
; OTHER INFORMATION: n = A,T,C or G
; US-10-779-543-5864

Query Match
Best Local Similarity 45.0%; Score 22; DB 6; Length 808;
Matches 28; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Cy 1 TATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGGC 41
Db 333 TATGATTTCTGCGCCCTGTGATNNNTACCATGAACGCG 373

RESULT 7
; Sequence 34311, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 34311
; LENGTH: 201
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-767-471-34311

Query Match
Best Local Similarity 45.0%; Score 21.6; DB 6; Length 201;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Cy 1 TATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGGCACA 44
Db 155 TTGATCAGTTTCATATTCCTGCTGGGTACTATGAAAAAGAAC 198
```

```

; Sequence 14458, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO: 14458
; LENGTH: 2895
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS8018_1
; US-10-767-701-14458

Query Match
Best Local Similarity 45.0%; Score 21.6; DB 6; Length 2895;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Cy 2 ATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGGCACA 45
Db 1033 ATAAGCAACGGGTGCGAGTTGTTGACCTCATGATTAATGCTCAT 1076

RESULT 9
; Sequence 10717, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; FILE REFERENCE: CLO01505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 10717
; LENGTH: 546025
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)...(546025)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
; US-10-767-471-10717

Query Match
Best Local Similarity 45.0%; Score 21.6; DB 6; Length 546025;
Matches 30; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Cy 1 TATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGGCACA 44
Db 329443 TTGATCAGTTTCATATTCCTGCTGGGTACTATGAAAAAGAAC 329486

RESULT 10
; Sequence 31393, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE REFERENCE: 38-21(53535)B
```

;; CURRENT APPLICATION NUMBER: US/10/767,701
;; CURRENT FILING DATE: 2004-01-29
;; NUMBER OF SEQ ID NOS: 63128
;; SEQ ID NO 31393
;; LENGTH: 607
;; TYPE: DNA
;; ORGANISM: Sorghum bicolor
;; FEATURE:
;; NAME/KEY: unsure
;; LOCATION: (1)..(607)
;; OTHER INFORMATION: unsure at all n locations
;; FEATURE:
US-10-767-701-31393

Query Match 44.6%; Score 21.4; DB 6; Length 607;
Best Local Similarity 66.0%; Pred. No. 9.5;
Matches 31; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 2 ATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGACATGCA 48
DB 306 AGATGCGCTGTGAGCTTTATTATGACACCATTAATGACCAACA 260

RESULT 11
US-10-765-790-41/c
;; Sequence 41, Application US/10765790
;; GENERAL INFORMATION:
;; APPLICANT: Bayer Healthcare LLC
;; APPLICANT: Beard, Chris
;; APPLICANT: Burgess, Chris
;; APPLICANT: Gannon, Allison
;; APPLICANT: Harvey, Jeanne
;; APPLICANT: Lechner, John F.
;; APPLICANT: Li, Zheng
;; TITLE OF INVENTION: Identification and Verification of Methylation Marker Sequences
;; FILE REFERENCE: 1657/2035
;; CURRENT APPLICATION NUMBER: US/10/765,790
;; CURRENT FILING DATE: 2004-01-27
;; PRIOR APPLICATION NUMBER: US 10/737,082
;; PRIOR FILING DATE: 2003-12-16
;; NUMBER OF SEQ ID NOS: 300
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 41
;; LENGTH: 26371
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-765-790-41

Query Match 43.8%; Score 21; DB 6; Length 26371;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 30; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 2 ATGATCAGCTGTTACCGGTGTTGTTACCCATGAATGACATG 46
DB 15911 AAGCCAGCTGTGATCTTGAATGCTCCCAACAATAGAAACATG 15867

RESULT 12
US-10-783-271-34
;; Sequence 34, Application US/10783271
;; GENERAL INFORMATION:
;; APPLICANT: Veridex, LLC
;; APPLICANT: Wang, Yixin
;; TITLE OF INVENTION: BREAST CANCER PROGNOSTICS
;; FILE REFERENCE: VDX-5003 USNP
;; CURRENT APPLICATION NUMBER: US/10/783,271
;; CURRENT FILING DATE: 2004-02-20
;; NUMBER OF SEQ ID NOS: 111
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO 34
;; LENGTH: 175737
;; TYPE: DNA

;; ORGANISM: human
US-10-783-271-34

Query Match 43.3%; Score 20.8; DB 6; Length 175737;
Best Local Similarity 70.0%; Pred. No. 58;
Matches 28; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 9 GCTGTTACCGGTGTTGTTACCCATGAATGACATGCA 48
DB 116051 GCTGCTATCTGGAGCTTTCCTCGAATGACACCTCCA 116090

RESULT 13
US-10-779-543-15761/c
;; Sequence 15761, Application US/10779543
;; GENERAL INFORMATION:
;; APPLICANT: Williams et al
;; TITLE OF INVENTION: GENE PRODUCTS DIFFERENTIALLY EXPRESSED
;; FILE REFERENCE: 2300-21302
;; CURRENT APPLICATION NUMBER: US/10/779,543
;; CURRENT FILING DATE: 2004-02-12
;; PRIOR APPLICATION NUMBER: 10/076,555
;; PRIOR FILING DATE: 2002-02-15
;; PRIOR APPLICATION NUMBER: 09/217,471
;; PRIOR FILING DATE: 1998-12-21
;; PRIOR APPLICATION NUMBER: 60/068,755
;; PRIOR FILING DATE: 1997-12-23
;; PRIOR APPLICATION NUMBER: 60/080,664
;; PRIOR FILING DATE: 1998-04-03
;; PRIOR APPLICATION NUMBER: 60/105,234
;; PRIOR FILING DATE: 1998-10-21
;; PRIOR APPLICATION NUMBER: 09/297,648
;; PRIOR FILING DATE: 2000-04-10
;; PRIOR APPLICATION NUMBER: PCT/US99/01619
;; PRIOR FILING DATE: 1999-01-28
;; PRIOR APPLICATION NUMBER: 60/072,910
;; PRIOR FILING DATE: 1998-01-28
;; PRIOR APPLICATION NUMBER: 60/075,954
;; PRIOR FILING DATE: 1998-02-24
;; PRIOR APPLICATION NUMBER: 60/080,114
;; PRIOR FILING DATE: 1998-03-31
;; Remaining Prior Application data removed - See file wrapper or PALM.
;; NUMBER OF SEQ ID NOS: 23767
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 15761
;; LENGTH: 421
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-779-543-15761

Query Match 42.9%; Score 20.6; DB 6; Length 421;
Best Local Similarity 74.3%; Pred. No. 19;
Matches 26; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 11 TGGTTACCGGTGTTGTTACCCATGAATGACAT 45
DB 225 TTGTTACTGTGCGATTCATCCCATGAATATTCACAT 191

RESULT 14
US-10-781-469-49
;; Sequence 49, Application US/10781469
;; GENERAL INFORMATION:
;; APPLICANT: Reddy, Koopa
;; APPLICANT: Guegler, Karl, J.
;; APPLICANT: Au-Young, Janice K.
;; TITLE OF INVENTION: COMPOSITION FOR DETECTION OF GENES ENCODING RECEPTOR PROTEINS
;; FILE REFERENCE: PA-0014-1 CON
;; CURRENT APPLICATION NUMBER: US/10/781,469
;; CURRENT FILING DATE: 2004-02-17
;; PRIOR APPLICATION NUMBER: US 09/496,321
;; PRIOR FILING DATE: 2000-02-01

PRIOR APPLICATION NUMBER: US 60/118,905
PRIOR FILING DATE: 1999-02-05
NUMBER OF SEQ ID NOS: 125
SOFTWARE: PERL Program
SEQ ID NO 49
LENGTH: 905
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID No: 2832568
US-10-781-469-49

Query Match 42.9%; Score 20.6; DB 6; Length 905;
Best Local Similarity 67.4%; Pred. No. 22;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 2 ATGATCAGCTGGTACCCGTTGTTACCATGAATGACACA 44
Db 605 ATAACTGGGTACACAGATCATAGCCCATGAATGACGCA 647

RESULT 15
US-10-775-169-52
Sequence 52, Application US/10775169
GENERAL INFORMATION:
APPLICANT: Wyeth
APPLICANT: Butczynski, Michael
APPLICANT: Twine, Natalie
APPLICANT: Dornier, Andrew
APPLICANT: Trepicchio, William
TITLE OF INVENTION: Method for Monitoring Drug Activities In Vivo
FILE REFERENCE: AM101080 (031896-015000)
CURRENT APPLICATION NUMBER: US/10/775,169
CURRENT FILING DATE: 2004-02-11
NUMBER OF SEQ ID NOS: 5278
SOFTWARE: PatentIn version 3.2
SEQ ID NO 52
LENGTH: 198161
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-169-52

Query Match 42.9%; Score 20.6; DB 6; Length 198161;
Best Local Similarity 67.4%; Pred. No. 72;
Matches 29; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

Qy 6 TCAGCTGTACCCGTTGTTACCATGAATGACACATGCA 48
Db 175379 TCACATGTGTCTCATGTTCTTAACACGAATGTCTCAGGGA 175421

Search completed: March 17, 2004, 08:25:38
Job time : 8.32948 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456
Sequence: 1 atgaaactttaagtcgc.....ccacgcctaacagctatata 456

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending Patents NA Main:

```

1: /cgn2_6/ptodata/2/pna/US0955_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US0956_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US0957_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US0958_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US0959_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US0960_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US0961_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US0962_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US0963_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US0964_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US0965_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US0966_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US0967_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US0968_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US0969_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US0970_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US0971_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US0972_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US0973_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US0974_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US0975_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US0976_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US0977_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US0978_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US0979_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US0980_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US0981_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US0982_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US0983_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US0984_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US0985_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US0986_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US0987_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US0988_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US0989_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US0990_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US0991_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US0992_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US0993_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US0994_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US0995_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US0996_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US0997_COMB.seq.*

```

```

44: /cgn2_6/ptodata/2/pna/US1008_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US1009_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US1010_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US1011_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US1012_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US1013_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US1014_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US1015_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US1016_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US1017_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US1018_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US1019_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US1020_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US1021_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US1022_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US1023_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US1024_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US1025_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US1026_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US1027_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US1028_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US1029_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US1030_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US1031_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US1032_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US1033_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US1034_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US1035_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US1036_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US1037_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US1038_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US1039_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US1040_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US1041_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US1042_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US1043_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US1044_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US1045_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US1046_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US1047_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US1048_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US1049_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US1050_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US1051_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US1052_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US1053_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US1054_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US1055_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US1056_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US1057_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US1058_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US1059_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US1060_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US1061_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US1062_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US1063_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US1064_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US1065_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US1066_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US1067_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US1068_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US1069_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US1070_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US1071_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US1072_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US1073_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US1074_COMB.seq.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23	US-09-543-407-11
2	404.8	88.8	456	6	US-08-233-642A-56
3	404.8	88.8	456	23	US-09-543-407-11
4	376	82.5	456	23	US-09-543-407-13
5	361.6	79.3	456	23	US-09-543-407-19
6	355.2	77.3	456	23	US-09-543-407-17
7	353.6	77.5	456	23	US-09-543-407-23
8	353.6	77.5	456	23	US-09-543-407-25
9	352	77.2	456	23	US-09-543-407-15
10	350.4	76.8	456	23	US-09-543-407-21
11	350.4	76.8	456	23	US-09-543-407-27
12	347.2	76.1	456	23	US-09-543-407-29
13	315.2	69.1	456	23	US-08-233-642A-54
14	222.4	48.8	477	17	US-09-252-691C-172
15	222.4	48.8	477	17	US-09-252-691C-172
16	222.4	48.8	477	51	US-10-417-886-172
17	216	47.4	456	23	US-09-543-407-3
18	216	47.4	456	13	US-08-978-878-3
19	216	47.4	456	31	US-09-741-873B-1
20	213.6	46.8	453	92	US-60-352-946-1
21	213.6	46.8	453	101	US-60-444-371-1
22	172	37.7	396	13	US-08-978-878-1
23	163.2	35.8	360	31	US-09-741-873B-1
24	52.2	11.4	78	23	US-09-543-407-53
25	49.4	10.8	78	23	US-09-543-407-52
26	48	10.5	78	23	US-09-543-407-9
27	40.8	8.9	78	23	US-09-543-407-48
28	40.8	8.9	100	45	US-10-146-492B-79
29	40.6	8.9	78	23	US-09-543-407-49
30	39.4	8.6	2000	36	US-09-887-272A-5263
31	39	8.6	100	45	US-10-146-492B-78
32	37.4	8.2	561	17	US-09-252-691C-171
33	37.4	8.2	561	17	US-09-252-691C-171
34	37.4	8.2	561	51	US-10-417-886-171
35	37.4	8.2	1304	51	US-10-437-963-1973
36	35.4	7.8	100	45	US-10-146-492B-77
37	35.4	7.8	3411	1	PCT-US01-08631-24330
38	35.4	7.8	3412	1	PCT-US01-08631-24948
39	34.6	7.6	456	23	US-09-543-407-4
40	34.2	7.5	385	18	US-09-333-535A-288
41	34.2	7.5	385	25	US-09-615-606A-909
42	34.2	7.5	396	64	US-60-089-524-4522
43	34.2	7.5	616	51	US-10-424-599-88945
44	34.2	7.5	880	27	US-09-654-617-2638
45	34.2	7.5	880	29	US-09-684-016-2638

ALIGNMENTS

RESULT 1

US-09-543-407-11

Sequence 11, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Collins, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR

TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES

FILE REFERENCE: 920043.406

CURRENT APPLICATION NUMBER: US/09/543.407

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11

LENGTH: 456

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga

OTHER INFORMATION: sequence containing the replacement fragment

OTHER INFORMATION: encoding P13 from GPe3 of *Leishmania* major.

US-09-543-407-11

Query Match 100.0%; Score 456; DB 23; Length 456;

Best Local Similarity 100.0%; Pred. No. 7.6e-136;

Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAATTTTAAATGAGATTCGACATGCTGTTCTGCGTCTGCT	60
DB	1	ATGAATTTTAAATGAGATTCGACATGCTGTTCTGCGTCTGCT	60
QY	61	GGGCTGTTTCAATGAGGCGCGCGGATTCATATACGCGCGCATATGTTCCGCG	120
DB	61	GGGCTGTTTCAATGAGGCGCGCGGATTCATATACGCGCGCATATGTTCCGCG	120
QY	121	CCGACTCAACGTTGACATTTATCATGACGTTCCGCTACGCTGCTCTGCA	180
DB	121	CCGACTCAACGTTGACATTTATCATGACGTTCCGCTACGCTGCTCTGCA	180
QY	181	AGGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTTACGCGCGCAT	240
DB	181	AGGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTTACGCGCGCAT	240
QY	241	GTAGGCGAGGTCGCGATATGATCTATGATGATGATGATGATGATGATGATGATGAT	300
DB	241	GTAGGCGAGGTCGCGATATGATCTATGATGATGATGATGATGATGATGATGATGAT	300
QY	301	GCACATGACCATGACATGACATGACATGACATGACATGACATGACATGACATGACAT	360
DB	301	GCACATGACCATGACATGACATGACATGACATGACATGACATGACATGACATGACAT	360
QY	361	AATTAAGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	420
DB	361	AATTAAGCGCGCTGTTATGATGATGATGATGATGATGATGATGATGATGATGATGAT	420
QY	421	GCACATGACCATGACATGACATGACATGACATGACATGACATGACATGACATGACAT	480
DB	421	GCACATGACCATGACATGACATGACATGACATGACATGACATGACATGACATGACAT	480

RESULT 2

US-08-233-642A-56

Sequence 56, Application US/08233642A

GENERAL INFORMATION:

APPLICANT: Kay, William W.

APPLICANT: Collins, S. Karen

APPLICANT: Clouthier, Sharon C.

APPLICANT: Doran, James L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

TITLE OF INVENTION: BASED VACCINES

NUMBER OF SEQUENCES: 58

NUMBER OF SEQUENCES: -

CORRESPONDENCE ADDRESS:

ADDRESS: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233.642A

FILING DATE: 26-APR-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.40C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 88.8%; Score 404.8; DB 6; Length 456;
Best Local Similarity 93.0%; Pred. No. 2.6e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATATACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATATACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
QY 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
QY 241 GTAGCCAGGGTCGGATTAATGTAATCTTAACTGACTCGAATGTTTAAAGAAATAT 300
DB 241 GTAGCCAGGGTCGGATTAATGTAATCTTAACTGACTCGAATGTTTAAAGAAATAT 300
QY 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTAATGTCGGCAATACGGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTAATGTCGGCAATACGGCGGT 360
QY 361 AATAACGCCGGCGGTGTTAATATGATCAGCTGTGTTACCGGTGTTTAAATG 420
DB 361 AATAACGCCGGCGGTGTTAATATGATCAGCTGTGTTACCGGTGTTTAAATG 420
QY 421 GCACATGCAACCAACGCCACGCTAACCAATATTA 456
DB 421 GGTTTGGCAACCAACGCCACGCTAACCAATATTA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1
Query Match 88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 93.0%; Pred. No. 2.6e-119;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATATACGGCGCGCAATAGTTCCGGC 120
DB 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATATACGGCGCGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCGGCTAACGCTTGCTCTGCAA 180
QY 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACCATTAACCAAGCGGTATGTGTAACGGCGCGAT 240
QY 241 GTAGCCAGGGTCGGATTAATGTAATCTTAACTGACTCGAATGTTTAAAGAAATAT 300
DB 241 GTAGCCAGGGTCGGATTAATGTAATCTTAACTGACTCGAATGTTTAAAGAAATAT 300
QY 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTAATGTCGGCAATACGGCGGT 360
DB 301 GCCACATCGACGATGGAACGCTAAACCTCCGATATTAATGTCGGCAATACGGCGGT 360
QY 361 AATAACGCCGGCGGTGTTAATATGATCAGCTGTGTTACCGGTGTTTAAATG 420
DB 361 AATAACGCCGGCGGTGTTAATATGATCAGCTGTGTTACCGGTGTTTAAATG 420
QY 421 GCACATGCAACCAACGCCACGCTAACCAATATTA 456
DB 421 GGTTTGGCAACCAACGCCACGCTAACCAATATTA 456

RESULT 4
US-09-543-407-13
Sequence 13, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding prf3 from GPe3 of Leishmania major.

US-09-543-407-13
Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 5.1e-110;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACCAATCGTAGTTTCTGGCAGTGTCTGGCT 60
QY 61 GGCGTCGTTCCACAATGGGGCGGGCGGTATCATATACGGCGCGCAATAGTTCCGGC 120

```

Db      61 GGCCTGTTCCACATGGGGCGCGCGGTATCTAAGCGCGGCAATAGTTCGGC 120
Qy      121 CCGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Db      121 CCGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Qy      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240
Db      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240
Qy      241 GTAGGCGGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
Db      241 GTAGGCGGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
Qy      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Db      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Qy      361 AATAACGCGCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db      361 CTGTTACCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy      421 GCACATGCAAAACAAGCCGCTAACCGATTTAA 456
Db      421 GGTTTGGCAACAAGCCGCTAACCGATTTAA 456

```

RESULT 5
US-09-543-407-19

```

; Sequence 19, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-19

```

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 2.3e-105;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

```

Qy      1 ATGAACTTTTAAAGTGGAGCATTCGACGATGTTCTGGAGAGTCTGGCT 60
Db      1 ATGAACTTTTAAAGTGGAGCATTCGACGATGTTCTGGAGAGTCTGGCT 60
Qy      61 GGGCTGTTCCACATGGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCGGC 120
Db      61 GGGCTGTTCCACATGGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCGGC 120
Qy      121 CCGGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Db      121 CCGGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Qy      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240
Db      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240

```

```

Qy      241 GTAGGCGGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
Db      241 GCGAGCGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
Qy      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Db      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Qy      361 AATAACGCGCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db      361 AATAACGCGCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy      421 GCACATGCAAAACAAGCCGCTAACCGATTTAA 456
Db      421 GGTTTGGCAACAAGCCGCTAACCGATTTAA 456

```

RESULT 6

US-09-543-407-17

```

; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

```

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.6e-103;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

Qy      1 ATGAACTTTTAAAGTGGAGCATTCGACGATGTTCTGGAGAGTCTGGCT 60
Db      1 ATGAACTTTTAAAGTGGAGCATTCGACGATGTTCTGGAGAGTCTGGCT 60
Qy      61 GGGCTGTTCCACATGGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCGGC 120
Db      61 GGGCTGTTCCACATGGGGCGCGCGGTATCATTAACGGCGGCAATAGTTCGGC 120
Qy      121 CCGGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Db      121 CCGGACTCAAGCTTGAGCATTTATGATGATGATGATGATGATGATGATGATGAT 180
Qy      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240
Db      181 AGCGATGCGCGGTAAATCTGAAACGACCATTCACCGAGCGGTATGATGATGATGAT 240
Qy      241 GTAGGCGGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
Db      241 GTAGGCGGAGGTGCGGATTAATGATGATGATGATGATGATGATGATGATGAT 300
Qy      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Db      301 GCCACATCGACAGTGAAGCGCTAAAACTCCGATATTACTGTCGCGCAATATGATCAG 360
Qy      361 AATAACGCGCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db      361 AATAACGCGCGGTGTTAATGATGATGATGATGATGATGATGATGATGATGATGAT 420

```

Qy 421 GCACATGCAACCAACGCGGCTAACGATTTAA 456
Db 421 GCTTTGGCAACCAACGCGGCTAACGATTTAA 456

RESULT 7

US-09-543-407-23

; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.6e-103;

Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCAATTCGACAACTGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCAATTCGACAACTGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTGTTCCACAAATGGGCGCGGCGGTATATCAATACGCGCGGCAATGTTCCGCG 120
Db 61 GGCCTGTTCCACAAATGGGCGCGGCGGTATATCAATACGCGCGGCAATGTTCCGCG 120
Qy 121 CCGGACTCAAGTGGAGCTTATCAAGTACGCTTCCGTAACGCTGCGCTCTGCAA 180
Db 121 CCGGACTCAAGTGGAGCTTATCAAGTACGCTTCCGTAACGCTGCGCTCTGCAA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTACGCGCGCAT 240
Db 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTACGCGCGCAT 240
Qy 241 GTAGCCGAGGTTGGGATTAATGTAATGTAAGTGAATGTTTCAAGAAATAT 300
Db 241 GTAGCCGAGGTTGGGATTAATGTAATGTAAGTGAATGTTTCAAGAAATAT 300
Qy 301 GCCACATCGACGAGTGAAGCTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCCACATCGACGAGTGAAGCTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy 361 AATAAGCCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTTACCAAGAAATG 420
Db 361 AATAAGCCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTTACCAAGAAATG 420
Qy 421 GCACATGCAACCAACGCGGCTAACGATTTAA 456
Db 421 GCTTTGGCAACCAACGCGGCTAACGATTTAA 456

RESULT 8

US-09-543-407-25

; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-25

Query Match 77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 8.6e-103;

Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCAATTCGACAACTGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCAATTCGACAACTGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTGTTCCACAAATGGGCGCGGCGGTATATCAATACGCGCGGCAATGTTCCGCG 120
Db 61 GGCCTGTTCCACAAATGGGCGCGGCGGTATATCAATACGCGCGGCAATGTTCCGCG 120
Qy 121 CCGGACTCAAGTGGAGCTTATCAAGTACGCTTCCGTAACGCTGCGCTCTGCAA 180
Db 121 CCGGACTCAAGTGGAGCTTATCAAGTACGCTTCCGTAACGCTGCGCTCTGCAA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTACGCGCGCAT 240
Db 181 AGCGATCCCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTACGCGCGCAT 240
Qy 241 GTAGCCGAGGTTGGGATTAATGTAATGTAAGTGAATGTTTCAAGAAATAT 300
Db 241 GTAGCCGAGGTTGGGATTAATGTAATGTAAGTGAATGTTTCAAGAAATAT 300
Qy 301 GCCACATCGACGAGTGAAGCTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCCACATCGACGAGTGAAGCTAAATCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy 361 AATAAGCCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTTACCAAGAAATG 420
Db 361 AATAAGCCGCGGCTGTTAATTAATGATCAGCTGTTACCGCTGTTTACCAAGAAATG 420
Qy 421 GCACATGCAACCAACGCGGCTAACGATTTAA 456
Db 421 GCTTTGGCAACCAACGCGGCTAACGATTTAA 456

RESULT 9

US-09-543-407-15

; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 2.8e-102;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
QY 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
DB 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACCACTTACCGAGCGGTTATGGTAAAGCGCCGAT 240
DB 181 AGCGATGCCCGTAAATCTGAAACGACCACTTACCGAGCGGTTATGGTAAAGCGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGATGACTGATGAGTCAAGATGGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGATGACTGATGAGTCAAGATGGTTTCAGAAATAT 300
QY 301 GCCACATCGACCAAGTGGAAAGCTTAAACCTCCGATTTACTGTCGCCCAATACGGCGGT 360
DB 301 GCCACATCGACCAAGTGGAAAGCTTAAACCTCCGATTTACTGTCGCCCAATACGGCGGT 360
QY 361 AATAACCGCGCGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 AATAACCGCGCGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GCACATGCAACCAACGCGCGCGTATCAACGATTTAA 456
DB 421 GGTTTGCAACCAACGCGCGCGTATCAACGATTTAA 456
```

RESULT 10
US-09-543-407-21
Sequence 21, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-21

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 9.3e-102;

Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
QY 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
DB 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACCACTTACCGAGCGGTTATGGTAAAGCGCCGAT 240
DB 181 AGCGATGCCCGTAAATCTGAAACGACCACTTACCGAGCGGTTATGGTAAAGCGCCGAT 240
QY 241 GTAGGCCAGGGTGGCGATTAATGATGACTGATGAGTCAAGATGGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGTGGCGATTAATGATGACTGATGAGTCAAGATGGTTTCAGAAATAT 300
QY 301 GCCACATCGACCAAGTGGAAAGCTTAAACCTCCGATTTACTGTCGCCCAATACGGCGGT 360
DB 301 GCCACATCGACCAAGTGGAAAGCTTAAACCTCCGATTTACTGTCGCCCAATACGGCGGT 360
QY 361 AATAACCGCGCGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
DB 361 AATAACCGCGCGTGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATG 420
QY 421 GCACATGCAACCAACGCGCGCGTATCAACGATTTAA 456
DB 421 GGTTTGCAACCAACGCGCGCGTATCAACGATTTAA 456
```

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-27

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 9.3e-102;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGCAGCATTCGACGAATCGTAGTTTCGAGAGCTCTGGCT 60
QY 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
DB 61 GCGCTGTTCCACAATGGGGCGGCGGCGTATCATACGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
DB 121 CCGGACTCAACGTTGAGCATTTATCATAGTGGTTCGCTAAAGCTGCTCTGCAA 180
```


Db 121 CCGAATCTAAGCTGAGATTATATAGTACGGTCCGTTACGCTGCGCTCTCTGAA 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACCAATTCACGAGCGGTTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACCAATTCACGAGCGGTTATGTAACGGCGCGAT 240
Qy 241 GTAGCGCGGCGGATTAATGTAATCTGTAAGTCAAGTATGTTTCAAGAAATAT 300
Db 241 TATATGCTGCTGTTACCGCGTGTATCCATGAAATGGCAATGATTCAGAAATAT 300
Qy 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAACCGCGCGTGTAAATGATGATCAGCTGTTACCGCGTGTATCCCATGAATG 420
Db 361 AATAACCGCGCGTGTAAATGATGATCAGCTGTTACCGCGTGTATCCCATGAATG 420
Qy 421 GCACATGCAACCAACGCGCAGGCTAACCAATTTAA 456
Db 421 GCTTTGGCAACCAACGCGCAGGCTAACCAATTTAA 456

RESULT 12

US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-29

Query Match 76.1%; Score 347.2; DB 23; Length 456;
Best Local Similarity 85.1%; Pred. No. 1e-100;
Matches 388; Conservative 0; Mismatches 68; Indels 0; Gaps 0;
Qy 1 ATGAACTTTTAAAGTGGAGCAATTCGAGCAATTCGAGTCTGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCAATTCGAGCAATTCGAGTCTGAGTCTGCT 60
Qy 61 GCGCTGCTTCAACAATGGGCGCGGCTAATCAATACGCGCGCAATGTTCCGCG 120
Db 61 GCGCTGCTTCAACAATGGGCGCGGCTAATCAATACGCGCGCAATGTTCCGCG 120
Qy 121 CCGAATCTAAGCTGAGATTATATAGTACGGTCCGTTACGCTGCGCTCTCTGAA 180
Db 121 CCGAATCTAAGCTGAGATTATATAGTACGGTCCGTTACGCTGCGCTCTCTGAA 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACCAATTCACGAGCGGTTATGTAACGGCGCGAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACCAATTCACGAGCGGTTATGTAACGGCGCGAT 240
Qy 241 GTAGCGCGGCGGATTAATGTAATCTGTAAGTCAAGTATGTTTCAAGAAATAT 300
Db 241 GTAGCGCGGCGGATTAATGTAATCTGTAAGTCAAGTATGTTTCAAGAAATAT 300

Qy 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Db 301 GCCACATCGACGACGTGGAACGCTAAACTCCGATATTAATGTCGGCAATACGGCGGT 360
Qy 361 AATAACCGCGCGTGTAAATGATGATCAGCTGTTACCGCGTGTATCCCATGAATG 420
Db 361 AATAACCGCGCGTGTAAATGATGATCAGCTGTTACCGCGTGTATCCCATGAATG 420
Qy 421 GCACATGCAACCAACGCGCAGGCTAACCAATTTAA 456
Db 421 GCTTTGGCAACCAACGCGCAGGCTAACCAATTTAA 456

RESULT 13

US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

Query Match 69.1%; Score 315.2; DB 6; Length 361;
Best Local Similarity 99.1%; Pred. No. 2e-90;
Matches 317; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 64 GTCGTTCAACAATGGGCGCGGCTAATCAATACGCGCGCAATAGTTCGGCGCG 123
Db 1 GTCGTTCAACAATGGGCGCGGCTAATCAATACGCGCGCAATAGTTCGGCGCG 60
Qy 124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACCGCTGCTCTGCAAGC 183
Db 61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACCGCTGCTCTGCAAGC 120

```
QY 184 GATGCCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTAGCGCCCGATGTA 243
|
|
|
Db 121 GATGCCCGTAATCTGAAGACCACTTACCGAGCGGTTATGTAGCGCCCGATGTA 180
|
|
|
QY 244 GGCAGCGGTGCGGATTAATGACTATTTGAAGTCACTGAGATGTTTCAAGAAATATGCC 303
|
|
|
Db 181 GGCAGCGGTGCGGATTAATGACTATTTGAAGTCACTGAGATGTTTCAAGAAATATGCC 240
|
|
|
QY 304 ACCATGACCACTGAGAACCTTAAATCTCCGATATTAATCTGTCGCGCAATACGCGGTAAT 363
|
|
|
Db 241 ACCATGACCACTGAGAACCTTAAATCTCCGATATTAATCTGTCGCGCAATACGCGGTAAT 300
|
|
|
QY 364 AACGCGCGCTGTTAATTA 383
|
|
|
Db 301 AACGCGCGCTGTTAATTA 320
|
|
|
```

RESULT 14

```
US-09-252-691-172
; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; PRIORITY FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172
```

```
Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
```

```
QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 60
|
|
|
Db 22 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 81
|
|
|
QY 61 GCGCTGTTCCACATGCGGCGCGGCGGTAATCATACGCGCGCAATAGTTCGCGC 120
|
|
|
Db 82 GCGCTGTTCCACATGCGGCGCGGCGGTAATCATACGCGCGCAATAGTTCGCGC 141
|
|
|
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGCTTCGCTGCTGCTGCTGCA 180
|
|
|
Db 142 CCGGACTCAACGTTGACATTTATCAGTACGCTTCGCTGCTGCTGCTGCTGCA 201
|
|
|
QY 181 AGCGATGCGCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 240
|
|
|
Db 202 AGCGATGCGCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 261
|
|
|
QY 241 GTAGGCGAGGTTGCGGATTAATGACTATTTGAATGAACTGACTGAGATGTTTCAAGAAATTAAT 300
|
|
|
Db 262 GTAGGCGAGGTTGCGGATTAATGACTATTTGAATGAACTGACTGAGATGTTTCAAGAAATTAAT 321
|
|
|
QY 301 GCCACATGACCACTGAGAACGCTAAATCTCCGATATTAATCTGCGCCCAATACGCGGCT 360
|
|
|
Db 322 GCCACATGACCACTGAGAACGCTAAATCTCCGATATTAATCTGCGCCCAATACGCGGCT 381
|
|
|
QY 361 AATTAAGCGCGGCGGTTAATTAATGATCACTGTTTACCGGTTGTTTACCCATGAATG 420
|
|
|
Db 382 GGCAGCGCGCGGCGGTTAGACGAGACGAGCGTCCGCTCAACGCTGACTGACACCGATT 441
|
|
|
QY 421 GCACATGCAACCAACGCGCAAGGCTAACCGATTA 456
|
|
|
Db 442 GCGTTTGGCAACCAACGCGCAAGGCTAACCGATTA 477
|
|
|
```

RESULT 15

US-09-252-691C-172

```
; Sequence 172, Application US/09252691C
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstein et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTER
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691C
; PRIORITY FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-07-24
; PRIOR APPLICATION NUMBER: US 60/074,787
; NUMBER OF SEQ ID NOS: 11326
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691C-172
```

```
Query Match 48.8%; Score 222.4; DB 17; Length 477;
Best Local Similarity 68.0%; Pred. No. 1.9e-60;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
```

```
QY 1 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 60
|
|
|
Db 22 ATGAACCTTTTAAAGTGGCAGCATTCGACGACATTCGAGTTTCTGCGAGTCTGGCT 81
|
|
|
QY 61 GCGCTGTTCCACATGCGGCGCGGCGGTAATCATACGCGCGCAATAGTTCGCGC 120
|
|
|
Db 82 GCGCTGTTCCACATGCGGCGCGGCGGTAATCATACGCGCGCAATAGTTCGCGC 141
|
|
|
QY 121 CCGGACTCAACGTTGACATTTATCAGTACGCTTCGCTGCTGCTGCTGCTGCA 180
|
|
|
Db 142 CCGGACTCAACGTTGACATTTATCAGTACGCTTCGCTGCTGCTGCTGCTGCTGCA 201
|
|
|
QY 181 AGCGATGCGCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 240
|
|
|
Db 202 AGCGATGCGCGTAAATCTGAAGACCACTTACCGAGCGGTTATGTTAGCGCGCGAT 261
|
|
|
QY 241 GTAGGCGAGGTTGCGGATTAATGACTATTTGAATGAACTGACTGAGATGTTTCAAGAAATTAAT 300
|
|
|
Db 262 GTAGGCGAGGTTGCGGATTAATGACTATTTGAATGAACTGACTGAGATGTTTCAAGAAATTAAT 321
|
|
|
QY 301 GCCACATGACCACTGAGAACGCTAAATCTCCGATATTAATCTGCGCCCAATACGCGGCT 360
|
|
|
Db 322 GCCACATGACCACTGAGAACGCTAAATCTCCGATATTAATCTGCGCCCAATACGCGGCT 381
|
|
|
QY 361 AATTAAGCGCGGCGGTTAATTAATGATCACTGTTTACCGGTTGTTTACCCATGAATG 420
|
|
|
Db 421 GCACATGCAACCAACGCGCAAGGCTAACCGATTA 456
|
|
|
Db 442 GCGTTTGGCAACCAACGCGCAAGGCTAACCGATTA 477
|
|
|
```

Search completed: March 18, 2004, 02:46:58
Job time : 2923.24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)

2406.048 Million cell updates/sec

Title: US-09-543-407-11

Perfect score: 456

Sequence: 1 atgaactcttaaaagtgc.....ccacgctaacagctatata 456

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA New:*

1: /cgn2_6/p/locdata/2/pna/PCT NEW COMB. seq.*

2: /cgn2_6/p/locdata/2/pna/US06 NEW COMB. seq.*

3: /cgn2_6/p/locdata/2/pna/US07 NEW COMB. seq.*

4: /cgn2_6/p/locdata/2/pna/US08 NEW COMB. seq.*

5: /cgn2_6/p/locdata/2/pna/US09 NEW COMB. seq.*

6: /cgn2_6/p/locdata/2/pna/US10 NEW COMB. seq.*

7: /cgn2_6/p/locdata/2/pna/US60 NEW COMB. seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216	47.4	675	5 US-09-741-873C-3	Sequence 3, Appli
2	172	37.7	396	5 US-09-741-873C-1	Sequence 1, Appli
3	33.4	7.3	3990	5 US-09-830-230A-629	Sequence 629, App
4	33.2	7.3	1355	6 US-10-045-674A-593	Sequence 593, App
5	32.4	7.1	150223	6 US-10-624-149A-1	Sequence 1, Appli
6	31.4	6.9	1588	6 US-10-767-701-14522	Sequence 14522, A
7	31.2	6.8	290892	6 US-10-767-471-10809	Sequence 10809, A
8	30.2	6.6	512	6 US-10-781-014-257	Sequence 257, App
9	30.2	6.6	1325	6 US-10-781-014-255	Sequence 255, App
10	29.6	6.5	24900	6 US-10-767-471-10736	Sequence 10736, A
11	29.4	6.4	300	6 US-10-767-701-28437	Sequence 28437, A
12	29.4	6.4	1400	6 US-09-545-213-2134	Sequence 2134, App
13	29.4	6.4	1400	7 US-60-545-213-2135	Sequence 2135, App
14	29.4	6.4	1400	7 US-60-545-213-6406	Sequence 6406, App
15	29.4	6.4	1400	7 US-60-545-213-6407	Sequence 6407, App
16	28.6	6.3	1186	1 PCT-US04-05654-1030	Sequence 1030, App
17	28.4	6.2	1350	6 US-10-784-480-291	Sequence 291, App
18	28.4	6.2	394468	6 US-60-548-091-5725	Sequence 5725, App
19	28.2	6.2	596	6 US-10-767-701-5378	Sequence 5378, App
20	28.2	6.2	610	6 US-10-767-701-8781	Sequence 8781, App
21	28.2	6.2	1746	1 PCT-US04-05654-2555	Sequence 2555, App
22	28	6.1	237	6 US-10-767-701-31417	Sequence 31417, App
23	28	6.1	600	7 US-60-545-213-1313	Sequence 1313, App
24	28	6.1	600	7 US-60-545-213-1314	Sequence 1314, App
25	28	6.1	600	7 US-60-545-213-5585	Sequence 5585, App
26	28	6.1	600	7 US-60-545-213-5586	Sequence 5586, App

C	27	28	6.1	2007	6	US-10-775-972-153	Sequence 153, App
C	28	28	6.1	2148	6	US-10-775-972-154	Sequence 154, App
C	29	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
C	30	27.8	6.1	1580	6	US-10-451-467A-239	Sequence 239, App
C	31	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
C	32	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
C	33	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, App
C	34	27.6	6.1	6224	6	US-10-453-372-1187	Sequence 1187, App
C	35	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, App
C	36	27.6	6.1	232882	6	US-10-767-471-10679	Sequence 10679, A
C	37	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, App
C	38	27.4	6.0	1295	6	US-10-100-683-5480	Sequence 5480, App
C	39	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
C	40	27.4	6.0	3416	6	US-10-100-683-3397	Sequence 3397, App
C	41	27.4	6.0	8206	6	US-10-021-698A-3529	Sequence 3529, App
C	42	27.4	6.0	186449	6	US-10-021-698A-715	Sequence 715, App
C	43	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, App
C	44	27.2	6.0	959	1	US-10-767-701-1446	Sequence 1446, App
C	45	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1	US-09-741-873C-3
Sequence 3, Application US/09741873C	
GENERAL INFORMATION:	
APPLICANT: Normark, Staffan	
APPLICANT: Olsen, Arne	
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation	
FILE REFERENCE: 012889-084	
CURRENT APPLICATION NUMBER: US/09/741,873C	
CURRENT FILING DATE: 2000-12-22	
PRIOR APPLICATION NUMBER: SE 8801723-1	
PRIOR FILING DATE: 1998-05-06	
PRIOR APPLICATION NUMBER: US 08/978,878	
PRIOR FILING DATE: 1997-11-26	
PRIOR APPLICATION NUMBER: US 07/347,189	
PRIOR FILING DATE: 1989-05-04	
PRIOR APPLICATION NUMBER: US 07/789,437	
PRIOR FILING DATE: 1991-11-06	
PRIOR APPLICATION NUMBER: US 07/970,846	
PRIOR FILING DATE: 1992-11-03	
PRIOR APPLICATION NUMBER: US 08/187,865	
PRIOR FILING DATE: 1994-01-28	
PRIOR APPLICATION NUMBER: US 08/318,519	
PRIOR FILING DATE: 1994-10-05	
NUMBER OF SEQ ID NOS: 11	
SOFTWARE: PatentIn version 3.0	
SEQ ID NO 3	
LENGTH: 675	
TYPE: DNA	
ORGANISM: Escherichia coli	
FEATURE:	
NAME/KEY: CDS	
LOCATION: (83)..(538)	
US-09-741-873C-3	
Query Match	47.4%; Score 216; DB 5; Length 675;
Best Local Similarity	67.1%; Pred. No. 5.7e-62;
Matches	306; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
QY	1 ATGAACCTTTAAAGTGCAGCATTCGCGCATCTTCCTGCGAGCTCTGCT 60
DB	83 ATGAACCTTTAAAGTGCAGCATTCGCGCATCTTCCTGCGAGCTCTGCT 142
QY	61 GCGGCTGCTTCAAGTGCAGCGCGCGCGGATATCATTAACGCGCGCATTAAGTCCGCG 120
DB	143 GCGGCTGCTTCAAGTGCAGCGCGCGCGGATATCATTAACGCGCGCATTAAGTCCGCG 202
QY	121 CCGGACTCAACGTTGAGCATTTATCATGATCGTTCCGCTAACGCTGCGCTTGCCTGCA 180

```

Db      203 CCAATTCTGACGTGAACATTTTACCAATGACGGTGGGTAAGTCTGACTTGGCTTCGCA 262
Qy      181 AGCGATGCCCGTAAATCTGAAAACGACATTAACCAAGACGGTATGTTAAGCGGCCGAT 240
Db      263 ACTGATGCCCGTAAATCTGACTTGAATTAACCAAGATGCGCGGTAATGTTGAGAT 322
Qy      241 GTAGCCAGGGGTGGGTAATATGATTAATGAACTGACTGAGATGTTTCAGAATATAT 300
Db      323 GTTGTGAGGGCTGAGTATGACGACTGATGATGATGACCAAGTGGCTTCGTAACAGC 382
Qy      301 GCCACATCGACCGATGTAACGCTAAATACTCCGATATTAATGTTGCGGCAATACGGCGGT 360
Db      383 GCTACTCTTATGATGATGTAACGCGAAATTTGGAATGACGTTAAACAGTTGCGTGGT 442
Qy      361 AATAACGCCCGCTGGTATTAATTAATGATGACGCTGTTACCCGTTGTTACCATGAATG 420
Db      443 GGCACACGGTGTGATGATTAACGACGACTGATTAATCTCCGTCACAGTACACAGTT 502
Qy      421 GCACATGCAACACGCGCACCGGTAACCATGATTA 456
Db      503 GCGCTTGTGTAACACGCGACCGCTCATGACTACTAA 538

```

```

RESULT 2
US-09-741-873C-1
/ Sequence 1, Application US/09741873C
/ GENERAL INFORMATION:
/ APPLICANT: Normark, Staffan
/ TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
/ FILE REFERENCE: 012889-084
/ CURRENT APPLICATION NUMBER: US/09/741,873C
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: SE 8801723-1
/ PRIOR FILING DATE: 1998-05-06
/ PRIOR APPLICATION NUMBER: US 08/978,878
/ PRIOR FILING DATE: 1997-11-26
/ PRIOR APPLICATION NUMBER: US 07/347,189
/ PRIOR FILING DATE: 1989-05-04
/ PRIOR APPLICATION NUMBER: US 07/789,437
/ PRIOR FILING DATE: 1991-11-06
/ PRIOR APPLICATION NUMBER: US 07/970,846
/ PRIOR FILING DATE: 1992-11-03
/ PRIOR APPLICATION NUMBER: US 08/187,865
/ PRIOR FILING DATE: 1994-01-28
/ PRIOR APPLICATION NUMBER: US 08/318,519
/ PRIOR FILING DATE: 1994-10-05
/ NUMBER OF SEQ ID NOS: 11
/ SOFTWARE: PatentIn version 3.0
/ SEQ ID NO 1
/ LENGTH: 396
/ TYPE: DNA
/ ORGANISM: Escherichia coli
US-09-741-873C-1

```

```

Query Match      37.7%; Score 172; DB 5; Length 396;
Best Local Similarity 64.6%; Pred. No. 1,7e-47;
Matches 256; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

```

```

Qy      61 GCGCTGTTCCACAATGGGCGCGCGGTAATCATTAACGCGCGGCAATAGTTCCGCG 120
Db      1 GGTGTGTCTCTGATGCGCGCGGTAACCAAGGTGGGTAATTAATAGCGGC 60
Qy      121 CCGGACTCAACGTTGACATTTATCAGTACGGTCCGTAACGCTGCGCTTCTTCGCA 180
Db      61 CCAAAATCTAGCTGAACATTTACGATACGCTGGGTAACCTGCACTTGTCTGCA 120
Qy      181 AGCGATGCCCGTAAATCTGAAAACGACATTAACCAAGACGGTATGTTAAGCGGCCGAT 240
Db      121 ACTGATGCCCGTAAATCTGACTTGAATTAACCAAGATGCGCGGTAATGTTGAGAT 180
Qy      241 GTAGCCAGGGGTGGGTAATATGATTAATGAACTGACTGAGATGTTTCAGAATATAT 300

```

```

Db      181 GTTGTGAGGGCTGATGATGACGACTCAATGCATGACCAAGCGCTTCGTTAACAGC 240
Qy      301 GCCACATCGACCGATGTAACGCTAAATACTCCGATATTAATGTTGCGGCAATACGGCGGT 360
Db      241 GCTACTCTTATGATGATGTAACGCGAAATTTGGAATGACGTTAAACAGTTGCGTGGT 300
Qy      361 AATAACGCCCGCTGGTATTAATTAATGATGACGCTGTTACCCGTTGTTTACCATGAATG 420
Db      301 GGCACACGGTGTGATGATTAACGACGACTGATTAATCTCCGTCACAGTACACAGTT 420
Qy      421 GCACATGCAACACGCGCACCGGTAACCATGATTA 456
Db      361 GCGCTTGTGTAACACGCGACCGCTCATGACTACTAA 396

```

```

RESULT 3
US-09-830-230A-629
/ Sequence 629, Application US/09830230A
/ GENERAL INFORMATION:
/ APPLICANT: Human Genome Sciences, Inc.
/ TITLE OF INVENTION: Lyme Disease Vaccines
/ FILE REFERENCE: PB481US
/ CURRENT APPLICATION NUMBER: US/09/830,230A
/ PRIOR FILING DATE: 2001-09-27
/ PRIOR APPLICATION NUMBER: PCT/US98/12718
/ PRIOR FILING DATE: 1998-06-18
/ PRIOR APPLICATION NUMBER: 60/057,483
/ PRIOR FILING DATE: 1997-09-03
/ PRIOR APPLICATION NUMBER: 60/053,344
/ PRIOR FILING DATE: 1997-07-22
/ PRIOR APPLICATION NUMBER: 60/053,377
/ PRIOR FILING DATE: 1997-07-22
/ PRIOR APPLICATION NUMBER: 60/050,359
/ PRIOR FILING DATE: 1997-06-20
/ NUMBER OF SEQ ID NOS: 756
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 629
/ LENGTH: 3990
/ TYPE: DNA
/ ORGANISM: Homo sapiens

```

```

/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1143)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1148)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1139)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1143)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1245)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1247)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1250)
/ OTHER INFORMATION: n equals a,t,g, or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (1251)
/ OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```

Query Match 7.3%; Score 33.4; DB 5; Length 3990;
Best Local Similarity 47.4%; Pred. No. 0.37; Indels 0; Gaps 0;
Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCGCTCTCTGCAAGCGATCCGCTTAATCTGAACGACCATTAACCG 216
Db GTTACGCTGCGCTCTCTGCAAGCGATCCGCTTAATCTGAACGACCATTAACCG 3492

Qy 217 AGCGTTATGTTACGCGCGCGATGTTAGCGCGGTGGATATAGTACTATTGAATG 276
Db ATTGCTGCTGCTATGTTAGCGCGGTATGATGATGCTGCGGATTTTGTAAAGATGAGATG 3552

Qy 277 ACTGAGATGCTTTCAGAAATTAATGCCACATCCAGCCAGTGAACGCTTAATCCCAT 336
Db AAGAGATGATTCAGATTCCTGCTCTTCTTTGAGGGGAGATGCTTAAGAGATGAGAA 3612

Qy 337 ATTACTGTCGCGCAATACGCGGCTAATAACG 367
Db TTGCTGTGAAGATGATGATGATGAGAAAG 3643

RESULT 4
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DYAX/002 CIP2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; PRIOR APPLICATION NUMBER: 2001-10-25
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; OTHER INFORMATION: nucleotide sequence
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1305)
US-10-045-674A-593

Query Match 7.3%; Score 33.2; DB 6; Length 1355;
Best Local Similarity 23.9%; Pred. No. 0.28;
Matches 83; Conservative 91; Mismatches 174; Indels 0; Gaps 0;

Qy 78 GGGGGGGGGCGGTATCATTAACGCGCGGCAATAGTTCCGCCCGCACTCAAGTTGAG 137
Db KGGTGGTGGTWSYGGYGGWGGYWSYGGYGGYGGYWSYGGYGGYWSYGGYGGWGG 827

Qy 138 CATTATCAGTACGCTTCCTGCAAGCGTCCGCTGCTCTGCAAGCGATGCCCGTAATC 197
Db YCGTWSYGGYGGYWSYGGYWSYGGYWSYGGYWSYGGYWSYGGYWSYGGYWSYGGYWSY 887

Qy 198 TGAACACCATTTACCAAGACGCTTATGTTAGTACGCGCGCATGTAGCCAGGTTGCGGA 257
Db YAAAGGGGCVATGACYGAAAYGCGYGAAYGAAYGAAAYGAAAYGAAAYGAAAYGAAAYG 947

Qy 258 TAATAGTACTATGACGATGACGATGCTGGAATGTTTCAGAAATATGCAACCATGACGATG 317
Db RTYGAATWSYGTCCCTACWGAATTAATGCTGCTGCTATGCAAGATGATGATGATGATG 1007

Qy 318 GAACGCTTAATAAATCCGATATTAATGCTGCGCCCAATACGCGGCTATTAACGCGGCTG 377
Db 1008 YGCTGCTGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1067

Qy 378 TAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425
Db 1068 YCARGTYGAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGYAGY 1115

RESULT 5
US-10-624-149A-1
; Sequence 1, Application US/10624149A
; GENERAL INFORMATION:
; APPLICANT: Neubauer, Antoine
; APPLICANT: Ziegler, Christina
; TITLE OF INVENTION: gm-Negative BHV-Mutants without Heterologous Elements
; FILE REFERENCE: 1/1372
; CURRENT APPLICATION NUMBER: US/10/624,149A
; PRIOR FILING DATE: 2003-07-21
; PRIOR APPLICATION NUMBER: 60/403,282
; PRIOR FILING DATE: 2002-08-14
; PRIOR APPLICATION NUMBER: DE 10233064
; PRIOR FILING DATE: 2002-07-19
; PRIOR APPLICATION NUMBER: DE 10317008
; PRIOR FILING DATE: 2003-04-11
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Version 3.0
; SEQ ID NO 1
; LENGTH: 150223
; TYPE: DNA
; ORGANISM: Equine herpesvirus 1
; PUBLICATION INFORMATION:
; AUTHORS: Telford, E.A.
; AUTHORS: Watson, M.S.
; AUTHORS: McBride, K.
; AUTHORS: Davison, A.J.
; TITLE: The DNA sequence of equine herpesvirus-1
; JOURNAL: Virology
; VOLUME: 189
; ISSUE: 1
; PAGES: 304-316
; DATE: JUL-1992
; DATABASE ACCESSION NUMBER: NC 001491, NCBI
; DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 3.5;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCGCAAGCGATGCGCGCTTAATCTGAACGACCATTAACCCAGAGCGTTATGTTAGC 232
Db 95542 CTGTGTAACGCGTGGGCTTATGTTATGATACGCTCATTTCCACAGCATGTTGCCCTACG 95601

Qy 233 GCGCGCATGATGACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 292
Db 95602 TCAAGCCCATGACGACGACGACGACGATGATGATGATGATGATGATGATGATGATG 95661

Qy 293 GAATTAATGCCACCATGACCATGACCATGACCATGACCATGACCATGACCATGACCATG 318
Db 95662 GAGAGCTTATTCACACGACGCTGTTG 95687

RESULT 6
US-10-767-701-14522
; Sequence 14522, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua

Db 234073 GTTGGTAATGATAGTGSTGTGCACGCGGTGATGCTAATGATTATGATGATGATGATGTT 234081

Db 315 GTACTGCTCAGCATGTCCTCGTCGTGAGAGTGTGAT 277

```
RESULT 9
US-10-781-014-255/c
; Sequence 255, Application US/10781014
; GENERAL INFORMATION:
; APPLICANT: Pompeju, Markus
; APPLICANT: Krogger, Burkhard
; APPLICANT: Schroder, Hartwig
; APPLICANT: Zelder, Oskar
; APPLICANT: Haberer, Gregor
; TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS
; TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY
; FILE REFERENCE: BGI-126PCN
; CURRENT APPLICATION NUMBER: US/10/781,014
; CURRENT FILING DATE: 2004-02-17
; PRIOR APPLICATION NUMBER: US 09/602,740
; PRIOR FILING DATE: 2000-06-23
; PRIOR APPLICATION NUMBER: 60/141,031
; PRIOR FILING DATE: 1999-06-25
; PRIOR APPLICATION NUMBER: 60/143,208
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: 60/151,572
; PRIOR FILING DATE: 1999-08-31
; PRIOR APPLICATION NUMBER: DE 19931412.8
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931413.6
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931419.5
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931420.9
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931424.1
; PRIOR FILING DATE: 1999-07-08
; PRIOR APPLICATION NUMBER: DE 19931428.4
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 784
; SEQ ID NO 255
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1303)
; OTHER INFORMATION: KX02596
US-10-781-014-255

Query Match          6.6%; Score 30.2; DB 6; Length 1326;
Best Local Similarity 56.6%; Pred. No. 2.7;
Matches 56; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

Qy 310 GACGAGTGAACGCTAAACCTCCGATTAATCTGCGGCAATACGGCGTAATACGCC 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1189 GCCAGGTGACCGCCGACACACACTTATTAAGACACCTCTTCAGACGACGAGCG 1130
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 370 GCGGTGTTATTAATGATCAGCTGTTACCCGTTGTT 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1129 GTAAGCTTTCAGACATGCTCGGTGCTGAGAGTGTAT 1091
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: RHEUMATOID ARTHRITIS, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: C0001505
; CURRENT APPLICATION NUMBER: US/10/767,471
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 50231
; SOFTWARE: FastSeq for Windows Version 4.0
```

```
; SEQ ID NO 10736
; LENGTH: 24900
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(24900)
; OTHER INFORMATION: n = A,T,C or G, or insertion/deletion polymorphism (see Tables 1-
US-10-767-471-10736

Query Match          6.5%; Score 29.6; DB 6; Length 24900;
Best Local Similarity 52.4%; Pred. No. 14;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 207 CATTACCCAGAGCGGTTATGTTACGCGCCGATGAGCCAGGTCGGATTAATAGTAC 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 10048 CAGCATCTGACTCTTGTAGTGTCTCAAAAGTCTCGGTCACTGTCATTAATGCTG 9989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 267 TATTGAATGCTGAGATGTTTCAGAAATTAATGCCACATGACGACGAGTGAACGCTAA 326
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 9988 TCCTGAACCCAACTGTAATGTAATTAATAATAGTCAATCTGTTGTTATATTAA 9929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 327 AAC 330
    |||
Db 9928 AATC 9925
    |||

RESULT 11
US-10-767-701-28437
; Sequence 28437, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: KOVALIC, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5335)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 28437
; LENGTH: 300
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1303)
; OTHER INFORMATION: KX02596
US-10-767-701-28437

Query Match          6.4%; Score 29.4; DB 6; Length 300;
Best Local Similarity 54.1%; Pred. No. 2.7;
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Qy 14 AAGTGACGAGATTGCGACGATCGTACTTCTGCGAGTGTCTGCGCTGCTCCAC 73
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 123 AAGAAAGCGGTACGAAGCCGCTGTTACGCTGCGCGCGGCGGTTACGCGGCTCCC 182
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 74 AATGGGCGCGCGGCGGTAATCATACGCGCGGCAATAGTTCCGCGCCG 124
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 183 GTAAGCGCGCGCGGCTTATGCGCGCGCGCGCTTACGCTGCGCAACCGG 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 12
US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; CURRENT FILING DATE: 2004-02-18
; NUMBER OF SEQ ID NOS: 303284
```

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2134

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-2134

Query Match 6.4%; Score 29.4; DB 7; Length 1400;

Best Local Similarity 45.2%; Pred. No. 5.1;

Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGCGCCGATGTAGCGGATGCGGATATAGTACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGGAGTTGGGGTATGCGGTTATTA 378

272 AACTGACTCAGAAATGTTTCAAGAAATATGCCACCATGACCAAGTGAACCGTAAACT 331

377 TTGGGGTTCCTCCATCCACTGCTTGAAAGCCAGAGACCATCTCTCACACTGGAAC 318

332 CCGATATTACTGTGCGCAATACGCGCGGTATTAACGCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGCAGCGCGCTCTGATGCAATTAACCTGCTGATCTGCTCATCAGGTTTCT 258

392 TGGTTACCGGTTGTTTACCCATGAAATGACATGCAACAGCCAGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAATGAAGAGGGGTGTGACCTG 199

RESULT 13

US-60-545-213-2135/c

Sequence 2135, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545, 213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 2135

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-2135

Query Match 6.4%; Score 29.4; DB 7; Length 1400;

Best Local Similarity 45.2%; Pred. No. 5.1;

Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGCGCGCATGTAGCCGCGGTGCGGATATAGTACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGGATGCGGTTGATGCGGTTATTA 378

272 AACTGACTCAGAAATGTTTCAAGAAATATGCCACCATGACCAAGTGAACCGTAAACT 331

377 TTGGGGTTCCTCCATCCACTGCTTGAAAGCCAGAGACCATCTCTCACACTGGAAC 318

332 CCGATATTACTGTGCGCAATACGCGCGGTATTAACGCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGCAGCGCGCTCTGATGCAATTAACCTGCTGATCTGCTCATCAGGTTTCT 258

392 TGGTTACCGGTTGTTTACCCATGAAATGACATGCAACAGCCAGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAATGAAGAGGGGTGTGACCTG 199

RESULT 14

US-60-545-213-6406/c

Sequence 6406, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545, 213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 6406

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-6406

Query Match 6.4%; Score 29.4; DB 7; Length 1400;

Best Local Similarity 45.2%; Pred. No. 5.1;

Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGCGCGCATGTAGCGGATGCGGATATAGTACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGGATGCGGTTGATGCGGTTATTA 378

272 AACTGACTCAGAAATGTTTCAAGAAATATGCCACCATGACCAAGTGAACCGTAAACT 331

377 TTGGGGTTCCTCCATCCACTGCTTGAAAGCCAGAGACCATCTCTCACACTGGAAC 318

332 CCGATATTACTGTGCGCAATACGCGCGGTATTAACGCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGCAGCGCGCTCTGATGCAATTAACCTGCTGATCTGCTCATCAGGTTTCT 258

392 TGGTTACCGGTTGTTTACCCATGAAATGACATGCAACAGCCAGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAATGAAGAGGGGTGTGACCTG 199

RESULT 15

US-60-545-213-6407/c

Sequence 6407, Application US/60545213

GENERAL INFORMATION:

APPLICANT: Wyeth

APPLICANT: Mounts, William Martin

TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug

FILE REFERENCE: AM101083 (031896-042099)

CURRENT APPLICATION NUMBER: US/60/545, 213

CURRENT FILING DATE: 2004-02-18

NUMBER OF SEQ ID NOS: 303284

SOFTWARE: PatentIn version 3.2

SEQ ID NO 6407

LENGTH: 1400

TYPE: DNA

ORGANISM: Homo sapiens

US-60-545-213-6407

Query Match 6.4%; Score 29.4; DB 7; Length 1400;

Best Local Similarity 45.2%; Pred. No. 5.1;

Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

212 CCCAGAGCGGTTATGTAACGCGCGCATGTAGCGGATGCGGATATAGTACTATTG 271

437 CCTGGGGGATAGCGCTTCAGATGAGCGGTGACCGGATGCGGTTGATGCGGTTATTA 378

272 AACTGACTCAGAAATGTTTCAAGAAATATGCCACCATGACCAAGTGAACCGTAAACT 331

377 TTGGGGTTCCTCCATCCACTGCTTGAAAGCCAGAGACCATCTCTCACACTGGAAC 318

332 CCGATATTACTGTGCGCAATACGCGCGGTATTAACGCGCGCTGTTATATGATCAGC 391

317 CCGTTGAGCAGCGCGCTCTGATGCAATTAACCTGCTGATCTGCTCATCAGGTTTCT 258

392 TGGTTACCGGTTGTTTACCCATGAAATGACATGCAACAGCCAGCGCTAACCG 450

257 GGGATCTCCCTCCAGTTGTTGTAATTTCTGAAATGAAGAGGGGTGTGACCTG 199

Thu Mar 18 12:27:56 2004

us-09-543-407-11.rnpn

Page 7

Search completed: March 17, 2004, 08:25:40
Job time : 71.6301 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaacttttaaaagtgc.....ccacgcttaaccagtactaa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3757730 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents NA Main:*

1: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
2: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
3: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
4: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
5: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
6: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
7: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
8: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
9: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
10: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
11: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
12: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
13: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
14: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
15: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
16: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
17: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
18: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
19: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
20: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
21: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
22: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
23: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
24: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
25: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
26: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
27: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
28: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
29: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
30: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
31: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
32: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
33: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
34: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
35: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
36: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
37: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
38: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
39: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
40: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
41: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
42: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*
43: /cgn2_6/prodata/2/pna/US0956_COMB.seq:*

44: /cgn2_6/prodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/prodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/prodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/prodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/prodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/prodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/prodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/prodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/prodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/prodata/2/pna/US105A_COMB.seq:*
54: /cgn2_6/prodata/2/pna/US105B_COMB.seq:*
55: /cgn2_6/prodata/2/pna/US106A_COMB.seq:*
56: /cgn2_6/prodata/2/pna/US106B_COMB.seq:*
57: /cgn2_6/prodata/2/pna/US107A_COMB.seq:*
58: /cgn2_6/prodata/2/pna/US107B_COMB.seq:*
59: /cgn2_6/prodata/2/pna/US108A_COMB.seq:*
60: /cgn2_6/prodata/2/pna/US108B_COMB.seq:*
61: /cgn2_6/prodata/2/pna/US109A_COMB.seq:*
62: /cgn2_6/prodata/2/pna/US109B_COMB.seq:*
63: /cgn2_6/prodata/2/pna/US110A_COMB.seq:*
64: /cgn2_6/prodata/2/pna/US110B_COMB.seq:*
65: /cgn2_6/prodata/2/pna/US111A_COMB.seq:*
66: /cgn2_6/prodata/2/pna/US111B_COMB.seq:*
67: /cgn2_6/prodata/2/pna/US112A_COMB.seq:*
68: /cgn2_6/prodata/2/pna/US112B_COMB.seq:*
69: /cgn2_6/prodata/2/pna/US113A_COMB.seq:*
70: /cgn2_6/prodata/2/pna/US113B_COMB.seq:*
71: /cgn2_6/prodata/2/pna/US114A_COMB.seq:*
72: /cgn2_6/prodata/2/pna/US114B_COMB.seq:*
73: /cgn2_6/prodata/2/pna/US115A_COMB.seq:*
74: /cgn2_6/prodata/2/pna/US115B_COMB.seq:*
75: /cgn2_6/prodata/2/pna/US116A_COMB.seq:*
76: /cgn2_6/prodata/2/pna/US116B_COMB.seq:*
77: /cgn2_6/prodata/2/pna/US117A_COMB.seq:*
78: /cgn2_6/prodata/2/pna/US117B_COMB.seq:*
79: /cgn2_6/prodata/2/pna/US118A_COMB.seq:*
80: /cgn2_6/prodata/2/pna/US118B_COMB.seq:*
81: /cgn2_6/prodata/2/pna/US119A_COMB.seq:*
82: /cgn2_6/prodata/2/pna/US119B_COMB.seq:*
83: /cgn2_6/prodata/2/pna/US120A_COMB.seq:*
84: /cgn2_6/prodata/2/pna/US120B_COMB.seq:*
85: /cgn2_6/prodata/2/pna/US121A_COMB.seq:*
86: /cgn2_6/prodata/2/pna/US121B_COMB.seq:*
87: /cgn2_6/prodata/2/pna/US122A_COMB.seq:*
88: /cgn2_6/prodata/2/pna/US122B_COMB.seq:*
89: /cgn2_6/prodata/2/pna/US123A_COMB.seq:*
90: /cgn2_6/prodata/2/pna/US123B_COMB.seq:*
91: /cgn2_6/prodata/2/pna/US124A_COMB.seq:*
92: /cgn2_6/prodata/2/pna/US124B_COMB.seq:*
93: /cgn2_6/prodata/2/pna/US125A_COMB.seq:*
94: /cgn2_6/prodata/2/pna/US125B_COMB.seq:*
95: /cgn2_6/prodata/2/pna/US126A_COMB.seq:*
96: /cgn2_6/prodata/2/pna/US126B_COMB.seq:*
97: /cgn2_6/prodata/2/pna/US127A_COMB.seq:*
98: /cgn2_6/prodata/2/pna/US127B_COMB.seq:*
99: /cgn2_6/prodata/2/pna/US128A_COMB.seq:*
100: /cgn2_6/prodata/2/pna/US128B_COMB.seq:*
101: /cgn2_6/prodata/2/pna/US129A_COMB.seq:*
102: /cgn2_6/prodata/2/pna/US129B_COMB.seq:*
103: /cgn2_6/prodata/2/pna/US130A_COMB.seq:*
104: /cgn2_6/prodata/2/pna/US130B_COMB.seq:*
105: /cgn2_6/prodata/2/pna/US131A_COMB.seq:*
106: /cgn2_6/prodata/2/pna/US131B_COMB.seq:*
107: /cgn2_6/prodata/2/pna/US132A_COMB.seq:*
108: /cgn2_6/prodata/2/pna/US132B_COMB.seq:*
109: /cgn2_6/prodata/2/pna/US133A_COMB.seq:*
110: /cgn2_6/prodata/2/pna/US133B_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

```

; FEATURE: INFORMATION: Recombinant Salmonella enteritidis 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match      100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 7,1e-135;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY      1 ATGAACCTTTTAAAGTGGCAGCACTTGGCAGCAATCGTAGTTTCTGGCAGTGTCTGCGT 60
Db      1 ATGAACCTTTTAAAGTGGCAGCACTTGGCAGCAATCGTAGTTTCTGGCAGTGTCTGCGT 60

QY      61 GGCCTCGTTCCACAAATGGGGCGGGCGGGTATCATAAACGGCGGGCAATAGTTCCGGC 120
Db      61 GGCCTCGTTCCACAAATGGGGCGGGCGGGTATCATAAACGGCGGGCAATAGTTCCGGC 120

QY      121 CCGGACTCAAAGTTGAGCAATTTATCAGTACGGTTCCGCTAAACGGTCGCGCTTCTGCAA 180
Db      121 CCGGACTCAAAGTTGAGCAATTTATCAGTACGGTTCCGCTAAACGGTCGCGCTTCTGCAA 180

QY      181 AGCGATCCCGGTAAATCTGAAACGACCATTAACCCAGAGCGGTTATGTGAACGGCGCGCAT 240
Db      181 AGCGATCCCGGTAAATCTGAAACGACCATTAACCCAGAGCGGTTATGTGAACGGCGCGCAT 240

QY      241 GTAGGCCAGGGGTGGGATTAATAGTACTATTGAACCTGAACCTGACAGATGGTTTCAGAAATAT 300
Db      241 GTAGGCCAGGGGTGGGATTAATAGTACTATTGAACCTGACAGATGGTTTCAGAAATAT 300

QY      301 GCCACCATGACCAAGTGGGAACGGCTAAACCTCCGATTTACTGTGCGGCATATGATCAG 360
Db      301 GCCACCATGACCAAGTGGGAACGGCTAAACCTCCGATTTACTGTGCGGCATATGATCAG 360

QY      361 CTGGTTACCCGTGTTGTTCACCATGAAATGGCAATGACCAAGCGTAAATGGTGTCAGGTT 420
Db      361 CTGGTTACCCGTGTTGTTCACCATGAAATGGCAATGACCAAGCGTAAATGGTGTCAGGTT 420

QY      421 GGTTTTGGCAACAACGCCACAGGCTTAACCAAGTATTTAA 456
Db      421 GGTTTTGGCAACAACGCCACAGGCTTAACCAAGTATTTAA 456

RESULT 2
US-08-233-642A-56
; Sequence 56, Application US/08233642A
; GENERAL INFORMATION:
; APPLICANT: Kay, William W.
; APPLICANT: Collinson, S. Karen

```

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match
Best Local Similarity 93.0%; Score 404.8; DB 6; Length 456;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

88.8%; Score 404.8; DB 6; Length 456;
Best Local Similarity 93.0%; Pred. No. 1.9e-118;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTCTGCAA 180
121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTCTGCAA 180
181 AGCGATGCCGTAAATCTGAAACGACCAATACCCAGAGCGGTATAGTAAACGCGCGCAT 240
181 AGCGATGCCGTAAATCTGAAACGACCAATACCCAGAGCGGTATAGTAAACGCGCGCAT 240
241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCTGACTGAGTAAATAT 300
241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCTGACTGAGTAAATAT 300
301 GCCACATTCGACGAGTGAACGCTAAATCTCGATATTAATCTGCGCCAAATATGATCAG 360
301 GCCACATTCGACGAGTGAACGCTAAATCTCGATATTAATCTGCGCCAAATATGATCAG 360
361 CTGGTACCGGTGTTGTTACCCATGAAATGGCAATGCAATGCGTATGCGTCAAGTT 420
361 AATAACCGCGCGTGTATATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
421 GGTTCGCAACAGCGCAGCGCTAACACGATTTAA 456
421 GGTTCGCAACAGCGCAGCGCTAACACGATTTAA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match
Best Local Similarity 88.8%; Score 404.8; DB 23; Length 456;
Matches 424; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

88.8%; Score 404.8; DB 23; Length 456;
Best Local Similarity 89.9%; Pred. No. 2.8e-111;
Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTCTGCAA 180
121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTTCCGCTAACGCTGCGCTCTGCAA 180
181 AGCGATGCCGTAAATCTGAAACGACCAATACCCAGAGCGGTATAGTAAACGCGCGCAT 240
181 AGCGATGCCGTAAATCTGAAACGACCAATACCCAGAGCGGTATAGTAAACGCGCGCAT 240
241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCTGACTGAGTAAATAT 300
241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGAATCTGACTGAGTAAATAT 300
301 GCCACATTCGACGAGTGAACGCTAAATCTCGATATTAATCTGCGCCAAATATGATCAG 360
301 GCCACATTCGACGAGTGAACGCTAAATCTCGATATTAATCTGCGCCAAATATGATCAG 360
361 CTGGTACCGGTGTTGTTACCCATGAAATGGCAATGCAATGCGTATGCGTCAAGTT 420
361 AATAACCGCGCGTGTATATCAGACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
421 GGTTCGCAACAGCGCAGCGCTAACACGATTTAA 456
421 GGTTCGCAACAGCGCAGCGCTAACACGATTTAA 456

RESULT 4
US-09-543-407-23
Sequence 23, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p13 from GPe3 of Leishmania major.

US-09-543-407-23

Query Match
Best Local Similarity 83.9%; Score 382.4; DB 23; Length 456;
Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

83.9%; Score 382.4; DB 23; Length 456;
Best Local Similarity 89.9%; Pred. No. 2.8e-111;
Matches 410; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGTAGTTCTGGCAGTCTTGCT 60
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120
61 GGCCTCGTTCCACAATGGGGGCGGCGGTATATCAATACGCGCGCAATAGTTCCGGC 120

Db 61 GGGCGTTCACAAATGGGCGCGCGGCTAATATACCGCGCGCAATAGTCCGGC 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Qy 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Qy 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Qy 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Db 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Qy 361 CTGGTTACCCGTTGTTTACCATGAATGCGACATGCAAGGTTATGGTGGCTGAGTT 420
Db 361 CATGAATGCGACATGCAATGCAATGCAAGGCTGATTCAGCGTAAATGGTGGCTGAGTT 420
Qy 421 GGTGGGCAACCAACCGCGCTAACCGATTTAA 456
Db 421 GGTGGGCAACCAACCGCGCTAACCGATTTAA 456

RESULT 5

US-09-543-407-11
Sequence 11, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-11

Query Match 82.5%; Score 376; DB 23; Length 456;
Best Local Similarity 89.0%; Pred. No. 3.2e-109;
Matches 406; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
Qy 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Db 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Qy 61 GGGCTGTTCCACAAATGGGCGCGCGGCTAATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGTTCCACAAATGGGCGCGCGGCTAATCATTAACGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Qy 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240

Qy 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Qy 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Db 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Qy 361 CTGGTTACCCGTTGTTTACCATGAATGCGACATGCAAGGTTATGGTGGCTGAGTT 420
Db 361 CATGAATGCGACATGCAATGCAATGCAAGGCTGATTCAGCGTAAATGGTGGCTGAGTT 420
Qy 421 GGTGGGCAACCAACCGCGCTAACCGATTTAA 456
Db 421 GGTGGGCAACCAACCGCGCTAACCGATTTAA 456

RESULT 6

US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania* major.
US-09-543-407-19

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 1.3e-104;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
Qy 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Db 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGTAAGTTCTGGAGTGTCTGGCT 60
Qy 61 GGGCTGTTCCACAAATGGGCGCGCGGCTAATCATTAACGGCGGCAATAGTTCCGGC 120
Db 61 GGGCTGTTCCACAAATGGGCGCGCGGCTAATCATTAACGGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Db 121 CCGGACTCAAGTTGAGCATTTATCACTAGCGTTCGCTAACGCTGGCTTGCTGCAA 180
Qy 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Db 181 AGCGATCCCGTAAATCTGAAAGCAACCATTAACCCAGCGGTTATGGTAAACGCGCCGAT 240
Qy 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGGATTAATAGTACTATTAAGTAACTGACTGAGATGGTTTCAGAAATAT 300
Qy 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Db 301 GCCACCATGACAGCAGTGAACCGCTAAATACTCCGATATTAAGTGGCCCAATATATGATG 360
Qy 361 CTGGTTACCCGTTGTTTACCATGAATGCGACATGCAAGGTTATGGTGGCTGAGTT 420
Db 361 CATGAATGCGACATGCAATGCAATGCAAGGCTGATTCAGCGTAAATGGTGGCTGAGTT 420

Qy 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456
Db 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456

RESULT 7

US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17

Query Match 77.9%; Score 355.2; DB 23; Length 456;

Best Local Similarity 86.2%; Pred. No. 1.5e-102; Indels 0; Gaps 0;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCGTTCCACATGAGGCGCGCGGTATATCAATACGCGCGGCAATAGTTCCGCG 120
Db 61 GGCCTCGTTCCACATGAGGCGCGCGGTATATCAATACGCGCGGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGATTTATCAAGTCCGTTCCGTTACGCTGCGCTTGTGAA 180
Db 121 CCGGACTCAACGTTGAGATTTATCAAGTCCGTTCCGTTACGCTGCGCTTGTGAA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCAATTCAGAGCGGTTATGTTACGCGCGCAT 240
Db 181 AGCGATCCCGTAAATCTGAAGACCAATTCAGAGCGGTTATGTTACGCGCGCAT 240
Qy 241 GTAGCCAGGCGGTATGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGCCAGGCGGTATGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACATTCGACAGTGAAGCTAAATCTCCATATTAATGTCGCGCAATATGATCAG 360
Db 301 GCCACATTCGACAGTGAAGCTAAATCTCCATATTAATGTCGCGCAATATGATCAG 360
Qy 361 CTGGTACCCGTTGTTTACCATGAAATGACATGCAATGCAATGCAATGCAATG 420
Db 361 AATAACGCGCGGTGTTTATCAAGCGCATCTGATTCACAGCTATATGTCGTCAGGTT 420
Qy 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456
Db 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456

RESULT 8

US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

Query Match 77.5%; Score 353.6; DB 23; Length 456;

Best Local Similarity 86.0%; Pred. No. 4.8e-102; Indels 0; Gaps 0;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

Qy 1 ATGAACCTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACCTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCGTTCCACATGAGGCGCGCGGTATATCAATACGCGCGGCAATAGTTCCGCG 120
Db 61 GGCCTCGTTCCACATGAGGCGCGCGGTATATCAATACGCGCGGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGATTTATCAAGTCCGTTCCGTTACGCTGCGCTTGTGAA 180
Db 121 CCGGACTCAACGTTGAGATTTATCAAGTCCGTTCCGTTACGCTGCGCTTGTGAA 180
Qy 181 AGCGATCCCGTAAATCTGAAGACCAATTCAGAGCGGTTATGTTACGCGCGCAT 240
Db 181 CTGGTACCCGTTGTTTACCATGAAATGACATGCAATGCAATGCAATGCAATG 240
Qy 241 GTAGCCAGGCGGTATGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Db 241 GTAGCCAGGCGGTATGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
Qy 301 GCCACATTCGACAGTGAAGCTAAATCTCCATATTAATGTCGCGCAATATGATCAG 360
Db 301 GCCACATTCGACAGTGAAGCTAAATCTCCATATTAATGTCGCGCAATATGATCAG 360
Qy 361 CTGGTACCCGTTGTTTACCATGAAATGACATGCAATGCAATGCAATGCAATG 420
Db 361 AATAACGCGCGGTGTTTATCAAGCGCATCTGATTCACAGCTATATGTCGTCAGGTT 420
Qy 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456
Db 421 GGTITGGCAACAGCGCCAGGCTTAACGATTTAA 456

RESULT 9

US-09-543-407-15
Sequence 15, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-15

Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 1.6e-101;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
   |||
DB 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
QY 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
   |||
DB 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
   |||
DB 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
QY 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
   |||
DB 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
   |||
DB 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
QY 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
   |||
DB 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
   |||
DB 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
   |||
DB 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
   |||
DB 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
QY 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
   |||
DB 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
QY 361 CTGGTTACCCGCTGTTTACCCATGAATGGCAGATCAGACGATGATGATGATGAT 420
   |||
DB 361 AATAACCCCGGCTGTTTACCCATGAATGGCAGATCAGACGATGATGATGATGAT 420
QY 421 GGTGGGCAACAGCGCGGCTAACCGATTTAA 456
   |||
DB 421 GGTGGGCAACAGCGCGGCTAACCGATTTAA 456
```

RESULT 10
US-09-543-407-21
Sequence 21, Application US/09543407

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 5.1e-101;

Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
   |||
DB 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
QY 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
   |||
DB 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
   |||
DB 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
QY 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
   |||
DB 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
QY 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
   |||
DB 181 AGCGATGCCCGTAAATCTGAAACGACCATTCACGAGCGGTTATGTAACGGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
   |||
DB 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
   |||
DB 241 GTAGGCCAGGGTGGGATTAATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
   |||
DB 301 GCCACCATGACGAGTGAAGCGTAAACCTCCGATTTACTGCGGCAATATGATCAG 360
QY 361 CTGGTTACCCGCTGTTTACCCATGAATGGCAGATCAGACGATGATGATGATGAT 420
   |||
DB 361 AATAACCCCGGCTGTTTACCCATGAATGGCAGATCAGACGATGATGATGATGAT 420
QY 421 GGTGGGCAACAGCGCGGCTAACCGATTTAA 456
   |||
DB 421 GGTGGGCAACAGCGCGGCTAACCGATTTAA 456
```

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match 76.8%; Score 350.4; DB 23; Length 456;
Best Local Similarity 85.5%; Pred. No. 5.1e-101;
Matches 390; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
   |||
DB 1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTCTGGAGTCTGCT 60
QY 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
   |||
DB 61 GGCCTGTTCCAAATGGGGGGGGGGGCGGTATCATTAACGGCGGCAATAGTCCGGC 120
QY 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
   |||
DB 121 CCGGACTCAAGTTGAGCATTTATCATGAGTTCGGTACGCTGCTGCTGCTGCTGCT 180
```



```

Db      121 CGGAGCTCAAGCTGAGATTATAGTACGGTCCGCTAACGGCCCTGCTCTGCA 180
Qy      181 AGCGATGCCCTTAATCTGAAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Db      181 AGCGATGCCCTTAATCTGAAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Qy      241 GTAGCCGAGGGTGGCGATTAATAGTACTATGAACTGACGAATGGTTTGAAGAAATAT 300
Db      241 TATGATCAGCTGTTTACCCGTTTACCGATGAAATGGACATGCAATTCAGAAATAT 300
Qy      301 GCCACATCGACGATGGAACGCTAAAACTCCGATATTAATGTCGCGCAATATATGATAG 360
Db      301 GCCACATCGACGATGGAACGCTAAAACTCCGATATTAATGTCGCGCAATATGATAG 360
Qy      361 CTGGTATCCCGTGTGTTTATCCCATGAAATGGACATGCAAGCTATGTCGTCAGGTT 420
Db      361 AATAACGGCGCGCTGTTTATCAGACCCCATCTGATTCAGCGTAAATGTCGTCAGGTT 420
Qy      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456
Db      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456

```

RESULT 12

US-09-543-407-29

Sequence 29, Application US/09543407

```

GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-29

```

Query Match 76.5%; Score 348.8; DB 23; Length 456;

Best Local Similarity 85.3%; Pred. No. 1.7e-100; Indels 0; Gaps 0;

Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

```

Qy      1 ATGAACTTTTAAAGTGGACAGATTCCGACCAATCGAGTTCTGCGAGTCTCTGCT 60
Db      1 ATGAACTTTTAAAGTGGACAGATTCCGACCAATCGAGTTCTGCGAGTCTCTGCT 60
Qy      61 GCGCTGTTTCAACAATGGGGCGCGCGGCTATCAATAACGGCGCGCAATAGTTCCGGC 120
Db      61 GCGCTGTTTCAACAATGGGGCGCGCGGCTATCAATAACGGCGCGCAATAGTTCCGGC 120
Qy      121 CGGAGCTCAAGCTGAGATTATAGTACGGTCCGCTAACGGCCCTGCTCTGCA 180
Db      121 CGGAGCTCAAGCTGAGATTATAGTACGGTCCGCTAACGGCCCTGCTCTGCA 180
Qy      181 AGCGATGCCCTTAATCTGAAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Db      181 AGCGATGCCCTTAATCTGAAAACGACATTAACCGAGCGGTATGTAAACGGCCGAT 240
Qy      241 GTAGCCGAGGGTGGCGATTAATAGTACTATGAACTGACGAATGGTTTGAAGAAATAT 300
Db      241 GTAGCCGAGGGTGGCGATTAATAGTACTATGAACTGACGAATGGTTTGAAGAAATAT 300

```

```

Qy      301 GCCACATCGACGATGGAACGCTAAAACTCCGATATTAATGTCGCGCAATATGATAG 360
Db      301 GCCACATCGATGATGCTGTTTACCCGTTTACCGATGAAATGGACATGCAAGCGCGT 360
Qy      361 CTGGTATCCCGTGTGTTTATCCCATGAAATGGACATGCAAGCTATGTCGTCAGGTT 420
Db      361 AATAACGGCGCGCTGTTTATCAGACCCCATCTGATTCAGCGTAAATGTCGTCAGGTT 420
Qy      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456
Db      421 GGTTTGGCAACAACGCCGCGCTAACGATTTAA 456

```

RESULT 13

US-08-233-642A-54

Sequence 54, Application US/08233642A

```

GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: -
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

```

Query Match 63.0%; Score 287.2; DB 6; Length 361;

Best Local Similarity 99.0%; Pred. No. 8.7e-81; Indels 0; Gaps 0;

Matches 289; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

Qy      64 GTCTTCACCAATGGGGCGCGCGGCTAATCATACGCGCGGCAATAGTTCCGGCCG 123
Db      1 GTCTTACCAACAGTGGGGCGCGCGGCTAATCATACGCGCGGCAATAGTTCCGGCCG 60
Qy      124 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCTTGTCTGCAAGC 183
Db      61 GACTCAACGTTGAGCATTTATCAGTACGTTCCGCTAACGCTGGCTTGTCTGCAAGC 120

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-13

Perfect score: 456

Sequence: 1 atgaacttttaaaagtgc.....ccacgctaaccgtatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 637860 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Pending Patente NA New:
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225.6	49.5	675	US-09-741-873C-3	Sequence 3, Appli
2	181.6	39.8	396	US-09-741-873C-1	Sequence 1, Appli
3	33	7.2	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-149A-1	Sequence 1, Appli
5	31.4	6.9	1588	US-10-767-701-14522	Sequence 14522, A
6	30.6	6.7	7096	US-10-021-698A-892	Sequence 892, App
7	30.6	6.7	196083	US-10-021-698A-707	Sequence 707, App
8	30.2	6.6	2007	US-10-775-972-153	Sequence 153, App
9	30.2	6.6	2148	US-10-775-972-154	Sequence 154, App
10	30	6.6	1355	US-10-045-674A-593	Sequence 593, App
11	29.6	6.5	2159	US-10-767-701-13730	Sequence 13730, A
12	29.6	6.5	24900	US-10-767-701-10736	Sequence 10736, A
13	29.6	6.5	318752	US-10-767-701-10553	Sequence 10553, A
14	29.4	6.4	1300	US-10-767-701-28437	Sequence 28437, A
15	29.2	6.4	1790242	US-10-767-701-10805	Sequence 10805, A
16	29	6.4	670	US-10-767-701-15726	Sequence 15726, A
17	29	6.4	861	US-10-767-701-3616	Sequence 3616, App
18	28.8	6.3	1857	PCT-US04-05654-2371	Sequence 2371, App
19	28.6	6.3	1186	PCT-US04-05654-1030	Sequence 1030, App
20	28.6	6.3	4634	US-10-791-799-10	Sequence 10, Appli
21	28.4	6.2	1543	US-10-767-701-1782	Sequence 1782, App
22	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, App
23	28.2	6.2	596	US-10-767-701-5378	Sequence 5378, App
24	28.2	6.2	610	US-10-767-701-8781	Sequence 8781, App
25	28.2	6.2	798	US-10-100-663-2192	Sequence 2192, App
26	28.2	6.2	1746	PCT-US04-05654-2555	Sequence 2555, App

27	28	6.1	237	6	US-10-767-701-31417	Sequence 31417, A
28	28	6.1	714	6	US-10-779-543-23469	Sequence 23469, A
29	27.8	6.1	390	5	US-09-796-592B-3647	Sequence 3647, App
30	27.8	6.1	749	6	US-10-767-701-12516	Sequence 12516, A
31	27.6	6.1	201	6	US-10-767-701-28293	Sequence 28293, A
32	27.6	6.1	402	5	US-09-969-034-1464	Sequence 1464, App
33	27.6	6.1	1052	6	US-10-767-701-16629	Sequence 16629, A
34	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
35	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, App
36	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, App
37	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, App
38	27.6	6.1	10251	6	US-10-045-674A-582	Sequence 582, App
39	27.6	6.1	232882	6	US-10-767-701-10679	Sequence 10679, A
40	27.4	6.0	819	6	US-10-417-884A-2244	Sequence 2244, App
41	27.4	6.0	823	6	US-10-767-701-20007	Sequence 20007, App
42	27.4	6.0	883	6	US-10-767-701-1287	Sequence 1287, App
43	27.4	6.0	2000	6	US-10-151-553-1	Sequence 1, Appli
44	27.2	6.0	543	1	PCT-US04-05654-1515	Sequence 1515, App
45	27.2	6.0	547	6	US-10-767-701-24545	Sequence 24545, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 880123-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match
Best Local Similarity 49.5%; Score 225.6; DB 5; Length 675;
Matches 312; Conservative 0; Mismatches 144; Indels 0; Gaps 0;
QY 1 ATGAACTTTTAAAGTGGCGACGATTTGGCGATCGTAGTTTGGCGAGTCTGGCT 60
DB 83 ATGAACTTTTAAAGTGGCGACGATTTGGCGATCGTAGTTTGGCGAGTCTGGCT 142
QY 61 GGCCTGCTTCAATGAGGCGGCGGATATCATTAACGCGGCGCAATAGTTCGGC 120
DB 143 GGTGTGTCTTCTGTCGCGGCGGCGGATATCATTAACGCGGCGCAATAGTTCGGC 202
QY 121 CCGACTCAACGTTGACATTTATCACTACGCTTCGCTAACGCTGCGTTCCTGCA 180

Db 203 CCAATTCTGAGCTGAACATTTACAGTACGGTGGGTAATCTTGCACTTGCTGCA 262
Qy 181 AGGATATGCCGTAATCTGAACACGACATTTACCGAGGGGTTATGTAACGCCCGAT 240
Db 263 ACTGATGCCCGTAATCTTGACTTGACTATTAACCAACATGGGGGGTATGATGAT 322
Qy 241 GTAGGACAGGGGTGGGTAATAGTATTAATTAATTAATTAATTAATTAATTAAT 300
Db 323 GTTGTGTCAGGGCTCAATGACAGCTCAATGATCTGAACCAAGTGGCTTCCGTAACGC 382
Qy 301 GCCACATGACAGCGTGAACGCTAAATTCGATATTAATGTCGGCCATATGATCAG 360
Db 383 GCTACTCTTGATCAGTGAACGCAAAATTCGAAATGACGTTAAACAGTTCCGTGT 442
Qy 361 CTGGTATCCCGTGTGTATCCCATGAAATGACATGCAAGGTAATGTCGTCAGGTT 420
Db 443 GGCACGCTGCTGCACTTACCAAGCTGATCTAATCTCCGTCGAACGTGACTAGGT 502
Qy 421 GGTTCGCAACACGCGGCTAACCGATTA 456
Db 503 GGGTTTGTAAACAGCGGACCGCTCATCTGACTAA 538

RESULT 2
US-09-741-873C-1

Sequence 1, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Olesen, Steffan
TITLE OF INVENTION: Fibrinectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 39.8%; Score 181.6; DB 5; Length 396;
Best Local Similarity 66.2%; Pred. No. 8,4e-52;
Matches 262; Conservative 0; Mismatches 134; Indels 0; Gaps 0;

Qy 61 GGCCTCGTCCCAATGGGCGGGGTAATCATTAACGCGCGGCAATATTTCCGCG 120
Db 1 GGTCTTCTCTCACTAGTACGGGGGGGTAACCAAGTGGTGGCGTAATATACCGGC 60
Qy 121 CCGACTCAACGTTGACATTTATCACTACGTTCCGTAACGCTCGCTGCTGCA 180
Db 61 CCAATTTCTAGCTGAACATTTACAGTACGTTGGGGTAACCTCTGCTGCA 120
Qy 181 AGGATGCCCGTAATCTGAACGACATTTACCAAGCGGTTATGTAACGCGCGAT 240
Db 121 ACTGATGCCCGTAATCTGACTTATTAACCAACATGGGGCGGTAATGTCGAGAT 180
Qy 241 GTAGGACAGGGTGGGTAATAGTATTAATTAATTAATTAATTAATTAATTAAT 300

Db 181 GTTGTGAGGGCTCAGATGACAGCTCAATGATCTGACCCAGCTGCTCGTAACAGC 240
Qy 301 GCCACATGACAGCGTGAACGCTAAATTCGATATTAATGTCGGCCATATGATCAG 360
Db 241 GCTACTCTTGATCAGTGAACGCAAAATTCGAAATGACGTTAAACAGTTCCGTGT 300
Qy 361 CTGGTATCCCGTGTGTATCCCATGAAATGACATGCAAGGTAATGTCGTCAGGTT 420
Db 301 GGCACGCTGCTGCACTTACCAAGCTGATCTAATCTCCGTCGAACGTGACTAGGT 360
Qy 421 GGTTCGCAACACGCGGCTAACCGATTA 456
Db 361 GGGTTTGTAAACAGCGGACCGCTCATCTGACTAA 396

RESULT 3
US-09-830-230A-629

Sequence 629, Application US/09830230A
GENERAL INFORMATION:
APPLICANT: Human Genome Sciences, Inc.
TITLE OF INVENTION: Lyme Disease Vaccines
FILE REFERENCE: PB481US
CURRENT APPLICATION NUMBER: US/09/830,230A
PRIOR FILING DATE: 2001-09-27
PRIOR APPLICATION NUMBER: PCT/US98/12718
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/057,483
PRIOR FILING DATE: 1997-09-03
PRIOR APPLICATION NUMBER: 60/053,344
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/053,377
PRIOR FILING DATE: 1997-07-22
PRIOR APPLICATION NUMBER: 60/050,359
PRIOR FILING DATE: 1997-06-20
NUMBER OF SEQ ID NOS: 756
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 629
LENGTH: 3990
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: (1135)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1143)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1146)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1210)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1247)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1250)
OTHER INFORMATION: n equals a,t,g, or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (1251)
OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

Query Match 7.2%; Score 33; DB 5; Length 3990;
Best Local Similarity 47.8%; Pred. No. 0.43;
Matches 96; Conservative 0; Mismatches 105; Indels 0; Gaps 0;

Qy 157 GCTAACGCTGCTGCTCTCTCAAAAGCGATGCCGTAATCTGAACGACCATTAACCG 216
Db 3433 GTTACGGCTGGGGCTGTGTGAGAGAGATGAGAGAGAGCTGACAGAGCTTACAAATCCG 3492

Qy 217 AGCGGTATGTGAACGCGCCGATGTAGGCCAGGGTGGATATAGTACTATTGAATG 276
Db 3493 ATTCTGCTCTATTTGGGAGAGGTAATGAGATGCTGGGATTTTGTGAAGATGAGATG 3552

Qy 277 ACTGAGATGGTTTCAGAAATATATGCGACCATTCAGACCGTGAAGCGTAAACCTCCGAT 336
Db 3553 AAGAGAGATGATCGAATGCTGCTGCTATTGCTTTGAGGGGAGATGCTAAGGATGAAAG 3612

Qy 337 ATTACTGTGCGCAATATGAT 357
Db 3613 TTGCTGTGAAGAGTAATGAT 3613

RESULT 4
US-10-624-149A-1
Sequence 1, Application US/10624149A
GENERAL INFORMATION:
APPLICANT: Neubauer, Antonie
TITLE OF INVENTION: GM-Negative EHV-Mutants without Heterologous Elements
FILE REFERENCE: 1/1372
CURRENT APPLICATION NUMBER: US/10/624,149A
CURRENT FILING DATE: 2003-07-21
PRIOR APPLICATION NUMBER: 60/403,282
PRIOR FILING DATE: 2002-08-14
PRIOR APPLICATION NUMBER: DE 10233064
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: DE 10317008
PRIOR FILING DATE: 2003-04-11
NUMBER OF SEQ ID NOS: 2
SOFTWARE: Patent Version 3.0
SEQ ID NO 1
LENGTH: 150223
TYPE: DNA
ORGANISM: Equine herpesvirus 1
PUBLICATION INFORMATION:
AUTHORS: Telford, E.A.
AUTHORS: Watson, M.S.
AUTHORS: McBride, K.
AUTHORS: Davison, A.J.
TITLE: The DNA sequence of equine herpesvirus-1
JOURNAL: Virology
VOLUME: 189
ISSUE: 1
PAGES: 304-316
DATE: JUL-1992
DATABASE ACCESSION NUMBER: NC 001491, NCBI
DATABASE ENTRY DATE: 2000-08-01
US-10-624-149A-1

Query Match 7.1%; Score 32.4; DB 6; Length 150223;
Best Local Similarity 51.4%; Pred. No. 3.6;
Matches 75; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Qy 173 CTCTGCAAGGATGCCGTAATCTGAAAGACCATTAACCGAGCGGTATGTAACG 232
Db 95542 CTGTGTGAACGCTGGGCTTGTGATACGGCTCAATTCACAGCATGTTGCTTAACG 95601

Qy 233 GCGCGCATGTAGCGCAGGTCGGATATAGTACTATTGAACCTGACGAATGTTTCA 292
Db 95602 TCAAGCCCATGAGCAGACCGGATATGTTTCTGTATACCACTCTCTGGGCGCGTCA 95661

Qy 293 GAATAATGCCCATGACCACTGCG 318

Db 95662 GAGAGCTTATTCACAGAGCTGTTG 95687

RESULT 5
US-10-767-701-14522
Sequence 14522, Application US/10767701
GENERAL INFORMATION:
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
FILE REFERENCE: 38-21(5353)B
CURRENT APPLICATION NUMBER: US/10/767,701
CURRENT FILING DATE: 2004-01-29
NUMBER OF SEQ ID NOS: 63128
SEQ ID NO 14522
LENGTH: 1588
TYPE: DNA
ORGANISM: Sorghum bicolor
FEATURE:
OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS2100_1
US-10-767-701-14522

Query Match 6.9%; Score 31.4; DB 6; Length 1588;
Best Local Similarity 44.5%; Pred. No. 0.99; Indels 0; Gaps 0;
Matches 125; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Qy 40 GTTTCGCAAGTCTCTGCTGCTGCGCTGCTTCACAAATGGGCGCGCGGTATCATAC 99
Db 455 GGTGTGGCGGCTTCGGTGGTGTGCTATGAGCGGTGGCGCGGTGCTATGTGTGCGC 514

Qy 100 GCGCGCGCAATATGTTCCGCGCGGACTCAACGTTTATGACATTTATGATAGCTTCCGCT 159
Db 515 GCGCGGTGCTATGTGTGTGTGTATGTGTGCGCAACTACGCGCAACAGGGTGGCGGTAT 574

Qy 160 AAGCTGCGCTGCTGCTGCAAGCGATGCCGTAATCTGAAAGACCATTAACCGAGC 219
Db 575 GGTGTGTGTGTGTGTATGTATGAGCTGACAGTGAAGTGAAGCACTTGTGCTGGA 634

Qy 220 GGTATGTAAAGCGCGCGATGTAGGCCAGGTCGGATATATAGTACTATTGAATGACT 279
Db 635 GTATGTAACAGCTTGTGACAGAGTAATCTGCTGCTACAGTGTGTTTGTGTAACCT 694

Qy 280 CAGATGCTTTCAGAAATATATCCACATGACCACTGAGCA 320
Db 695 GCTGTAGCTTGTGCGGTGTGCTTCCACTGCGCGCGCAAGGA 735

RESULT 6
US-10-021-698A-892
Sequence 892, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN BERDRECH, PAUL
APPLICANT: DUPUIS, JOSEPH
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044051
CURRENT APPLICATION NUMBER: US/10/021,698A
CURRENT FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
PRIOR FILING DATE: 2000-06-14
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: Patent 2.1
SEQ ID NO 892
LENGTH: 7096
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (7085)..(7087)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-892

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 6; Length 7096;
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 278 CTCGAGATGTTTCAGAAATTAATGCCACATGACGACGGAAGCTAAATCTCCGATA 337
DB 2964 CTCCTCAAGGGCCAGGAAACATCTTAAGATTAAGTCACTGGCCATTAACACCTTA 3023
QY 338 TTACTGTGGCCAAATATGATCAGCTGTGTTACCCGTGTGTTACCATGAAATGCGACATG 397
DB 3024 TTCAGGAAATTAATCTCTCTGCAAGGCGACCTGTGTGTGAAGCTGAGACTGTGAGT 3083
QY 398 CAAGCTAATGCTGCTCAGCTGTTGTTTGGCAACAGCCGCTAACCG 450
DB 3084 TCAGCAACGAGATCGGCCAAGTGGGGCTGCACTAAACCCCTGCTTACCG 3136

RESULT 7
US-10-021-698A-707

Sequence 707, Application US/10021698A
GENERAL INFORMATION:
APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSE
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
APPLICANT: PANDIT, SUNIL
TITLE OF INVENTION: NOCIBOTITE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021.698A
PRIOR FILING DATE: 2001-10-22
PRIORITY APPLICATION NUMBER: 60/211,749
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO: 707
LENGTH: 196083
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (236)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (89496)..(8955)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (141214)..(141313)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (182468)..(182567)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (193506)..(193605)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (194799)..(194898)
OTHER INFORMATION: a, t, c or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (195354)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (195460)
OTHER INFORMATION: a, t, c or g
FEATURE:
NAME/KEY: modified_base
LOCATION: (195648)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-707

Query Match
Best Local Similarity 6.7%; Score 30.6; DB 6; Length 196083;
Matches 84; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 278 CTCGAGATGTTTCAGAAATTAATGCCACATGACGACGGAAGCTAAATCTCCGATA 337
DB 45914 CTCCTCAAGGGCCAGGAAACATCTTAAGATTAAGTCACTGGCCATTAACACCTTA 45973
QY 338 TTACTGTGGCCAAATATGATCAGCTGTGTTACCCGTGTGTTACCATGAAATGCGACATG 397
DB 45974 TTCAGGAAATTAATCTCTCTGCAAGGCGACCTGTGTGTGAAGCTGAGACTGTGAGT 46033
QY 398 CAAGCTAATGCTGCTCAGCTGTTGTTTGGCAACAGCCGCTAACCG 450
DB 46034 TCAGCAACGAGATCGGCCAAGTGGGGCTGCACTAAACCCCTGCTTACCG 46086

RESULT 8
US-10-775-972-153/c

Sequence 153, Application US/10775972
GENERAL INFORMATION:
APPLICANT: Henderson, Robert A.
APPLICANT: Bangur, Chaitanya S.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.455C21
CURRENT APPLICATION NUMBER: US/10/775.972
PRIOR FILING DATE: 2004-02-10
NUMBER OF SEQ ID NOS: 563
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO: 153
LENGTH: 2007
TYPE: DNA
ORGANISM: Homo sapiens
US-10-775-972-153

Query Match
Best Local Similarity 6.6%; Score 30.2; DB 6; Length 2007;
Matches 104; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

QY 140 TTATCAGTACGCTGCTGCTTAAGCGCTGCTGCTGCAACGATGCCGTAATCTG 199
DB 1416 TTCCCAAGAAATTTCTTAAGAAATGTTTATTTTAAAGTAAAGTAAAGTT 1357
QY 200 AAACGACATTAACCCAGAGCGTTATGTTAAGCGCCGCTGTAAGCCAGGGTGGGATA 259
DB 1356 CACTGGGAATTTTGCAATGCAATGTAAGAAACTGAGAGCTGCCCTAGGGGAATTTACGATT 1297
QY 260 ATAGTACTATTAAGTCACTCAGAAATGTTTCAAAATATATGCCACCATGACCACTGGA 319
DB 1296 AGGCTTGAATTAATCACTCAGAAATGTTTCAAAATATATGCCCTCCATTAGTGT 1237
QY 320 ACCTAAATCTCCGATATTAATGCTGCGCAATATGATAGCGTGT 366
DB 1236 AGGAGAGAGCTCAAGTTCAGAGCTTCTCAITTTTACCAACAGTT 1190

RESULT 9

US-10-775-972-154/c
; Sequence 154, Application US/10775972
; GENERAL INFORMATION:
; APPLICANT: Henderson, Robert A.
; APPLICANT: Bangur, Chaitanya S.
; APPLICANT: Wang, Tonglong
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.455C21
; CURRENT APPLICATION NUMBER: US/10/775,972
; NUMBER OF SEQ ID NOS: 563
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 154
; LENGTH: 2148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-775-972-154

Query Match 6.6%; Score 30.2; DB 6; Length 2148;
Best Local Similarity 45.8%; Pred. No. 2.9; Indels 0; Gaps 0;
Matches 104; Conservative 0; Mismatches 123; Indels 0; Gaps 0;

Qy 140 TTATGAGTACGCTTCCGCTTAACGCTGCTGCTGCAAAAGCATCCCGTAATCTG 199
Db 1549 TTCCCGAGAAATCTCTACAAAGATGTTCAATTTTAAAGTAAAGTACTTAAAGTT 1490

Qy 200 AAACGACCATTAACCCAGAGCGGTTATGTAACGCGCCGATGAGCCAGGCGGATA 259
Db 1489 CACTGGGAATTTTGCATAGCAGTGAAGAACTGGAGCTGCTGAGGGAATTTTCAGCAT 1430

Qy 260 ATAGTACTATGATGCTGCTGAGAAATGTTTCAGAAATTAATGCAACCATTCAGCATGGA 319
Db 1429 AGGCTTGAATTAATCTACCCAGAGAAATCCAGTAATTAATCTGCTCCATTAATGAGT 1370

Qy 320 ACGCTAAAACTCCGATATTAATGTCGCGCAATATGATCAGCTGGTT 366
Db 1369 AGGAGAGAGCTCCAGATTCGAGAGCTTCTCATTTTACCAAAACAGTT 1323

RESULT 10
US-10-045-674A-593
; Sequence 593, Application US/10045674A
; GENERAL INFORMATION:
; APPLICANT: LADNER, ROBERT C.
; APPLICANT: COHEN, EDWARD H.
; APPLICANT: NASTRI, HORACIO G.
; APPLICANT: ROOKEY, KRISTIN L.
; APPLICANT: HOET, RENE
; APPLICANT: HOOGENDOORN, HENDRICUS R. J. M.
; TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
; TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
; TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
; FILE REFERENCE: DVAX/002.C1P2
; CURRENT APPLICATION NUMBER: US/10/045,674A
; CURRENT FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: 06/198,069
; PRIOR FILING DATE: 2000-04-17
; PRIOR APPLICATION NUMBER: 09/837,306
; NUMBER OF SEQ ID NOS: 635
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 593
; LENGTH: 1355
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: M13-III
; NAME/KEY: CDS
; LOCATION: (1)...(1305)

US-10-045-674A-593

Query Match 6.6%; Score 30; DB 6; Length 1355;
Best Local Similarity 25.1%; Pred. No. 2.7;
Matches 84; Conservative 80; Mismatches 170; Indels 0; Gaps 0;

Qy 57 GGCCTGGCTGCTTCCACAAATGGGGCGGCGGATATCATATACGCGCGGCATATGTC 116
Db 768 KGGTGGTGGTWSYGGYGGWGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGGYGG 827

Qy 117 CGGCCGAGACTCAAGTGTGACATTTATACGATGCTTCCGCTTACGCTGCTGCTCT 176
Db 828 YGGYWSYGGYGGYGGYWSYGGYWSYGGYATTTGATATYARABARATGCMAYGCTAA 887

Qy 177 GCAAAGGATGCCCGTAATCTGAAACGACCATTAACCCAGAGCGGTTATGTTAAACGGCC 236
Db 888 YAARGSGCYATGACYCARAAYGCVYAGAAAYGCRCTRCAWSTGAYGCAARGGYAA 947

Qy 237 CGATGTAGGCGCAGGTCGGATATATAGTACTATTAAGTGAAGTGAAGTTCAGAA 296
Db 948 RYTGAYWSYGTGCTGACGAYTAATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1007

Qy 297 TAATGCCACCATCGACCATGAGGAAAGCTAAATCCGATATTAATGCTGCGCAATATGA 356
Db 1008 YGGYCTGCTAAYGAYAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1067

Qy 357 TCAGCTGTTACCCGCTGTTTACCATGAATG 390
Db 1068 YCARGTGGWAGYAGKAYAAWMSCKYTWATG 1101

RESULT 11
US-10-767-701-13730
; Sequence 13730, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 13730
; LENGTH: 2159
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS49140_1
US-10-767-701-13730

Query Match 6.5%; Score 29.6; DB 6; Length 2159;
Best Local Similarity 48.8%; Pred. No. 4.6;
Matches 80; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

Qy 288 TTTCAGAAATPAATGCGACCATTCAGACGAGTGAACGCTTAATACTCCGATATTAATCTGCG 347
Db 682 TTACACTATTCATGACACCGAGACACGAGAACTGCGCAAAAAGATGATCAAGTTATGA 741

Qy 348 CCAATATGATCAGCTGTTTACCGTGTGTTTACCCAGGAATGCAATGCAATGCAATGCAAT 407
Db 742 CCAAGAGGAGAGCAAGCCCTGTGTGATTTCTCAATTTTGAAGCAATGTCGCGAGAT 801

Qy 408 GGTGCTCAGGTTGTTTGGCAACAAGCCAGCGCTTAACAGT 451
Db 802 GCGTGTGATGTTAGATTCACCAATGCTACAGAAATATGT 845

RESULT 12
US-10-767-471-10736/c
; Sequence 10736, Application US/10767471
; GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456

Sequence: 1 atgaactcttaaaagtcgc.....ccacgctaaccagtattaa 456

Scoring table:

IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending_Patents_NA_Main:

```
1: /cgn2_6/ptodata/2/pna/US06_PCTUS_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/2/pna/US092A_COMB.seq:*
17: /cgn2_6/ptodata/2/pna/US092B_COMB.seq:*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq:*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq:*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
25: /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
26: /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
27: /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
28: /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
29: /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
30: /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
31: /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
32: /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
33: /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
34: /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
35: /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
36: /cgn2_6/ptodata/2/pna/US098D_COMB.seq:*
37: /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
38: /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
39: /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
40: /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq:*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq:*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
```

```
44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq:*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq:*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq:*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq:*
53: /cgn2_6/ptodata/2/pna/US105_COMB.seq:*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq:*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq:*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
79: /cgn2_6/ptodata/2/pna/US6033A_COMB.seq:*
80: /cgn2_6/ptodata/2/pna/US6033B_COMB.seq:*
81: /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
82: /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
83: /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
84: /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
85: /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
86: /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
87: /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
88: /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
89: /cgn2_6/ptodata/2/pna/US6042_COMB.seq:*
90: /cgn2_6/ptodata/2/pna/US6043_COMB.seq:*
91: /cgn2_6/ptodata/2/pna/US6044_COMB.seq:*
92: /cgn2_6/ptodata/2/pna/US6045_COMB.seq:*
93: /cgn2_6/ptodata/2/pna/US6046_COMB.seq:*
94: /cgn2_6/ptodata/2/pna/US6047_COMB.seq:*
95: /cgn2_6/ptodata/2/pna/US6048_COMB.seq:*
96: /cgn2_6/ptodata/2/pna/US6049_COMB.seq:*
97: /cgn2_6/ptodata/2/pna/US6050_COMB.seq:*
98: /cgn2_6/ptodata/2/pna/US6051_COMB.seq:*
99: /cgn2_6/ptodata/2/pna/US6052_COMB.seq:*
100: /cgn2_6/ptodata/2/pna/US6053_COMB.seq:*
101: /cgn2_6/ptodata/2/pna/US6054_COMB.seq:*
102: /cgn2_6/ptodata/2/pna/US6055_COMB.seq:*
103: /cgn2_6/ptodata/2/pna/US6056_COMB.seq:*
104: /cgn2_6/ptodata/2/pna/US6057_COMB.seq:*
105: /cgn2_6/ptodata/2/pna/US6058_COMB.seq:*
106: /cgn2_6/ptodata/2/pna/US6059_COMB.seq:*
107: /cgn2_6/ptodata/2/pna/US6060_COMB.seq:*
108: /cgn2_6/ptodata/2/pna/US6061_COMB.seq:*
109: /cgn2_6/ptodata/2/pna/US6062_COMB.seq:*
110: /cgn2_6/ptodata/2/pna/US6063_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Description
Sequence 15, Appl
Sequence 56, Appl
Sequence 1, Appl
Sequence 19, Appl
Sequence 17, Appl
Sequence 11, Appl
Sequence 23, Appl
Sequence 25, Appl
Sequence 21, Appl
Sequence 27, Appl
Sequence 29, Appl
Sequence 54, Appl
Sequence 3, Appl
Sequence 172, Appl
Sequence 172, Appl
Sequence 172, Appl
Sequence 3, Appl
Sequence 3, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 170, Appl
Sequence 170, Appl
Sequence 170, Appl
Sequence 170, Appl
Sequence 72, Appl
Sequence 52, Appl
Sequence 9, Appl
Sequence 53, Appl
Sequence 78, Appl
Sequence 48, Appl
Sequence 79, Appl
Sequence 49, Appl
Sequence 171, Appl
Sequence 171, Appl
Sequence 6264, Appl
Sequence 6264, Appl
Sequence 6264, Appl
Sequence 20533, A
Sequence 24330, A
Sequence 29448, A
Sequence 3383, A
Sequence 3383, A

ALIGNMENTS

```

RESULT 1
US-09-543-407-15
: Sequence 15: Application US/09543407
: GENERAL INFORMATION:
: APPLICANT: White, Aaron P.
: APPLICANT: Doran, James L.
: APPLICANT: Collinson, S. Karen
: APPLICANT: Kay, William W.
: TITLE OF INVENTION: BACTERIAL FILTRAL SYSTEM FOR
: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
: FILE REFERENCE: 920043.406
: CURRENT APPLICATION NUMBER: US/09/543.407
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: FaastSeq for Windows Version 4.0
: SEQ ID NO 15
: LENGTH: 456
: TYPE: DNA
: ORGANISM: Artificial Sequence

```

```

; FEATURE: Recombinant Salmonella enteritidis 3b afaA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-15

```

Query Match	100.0%	Score 456;	DB 23;	Length 456;
Best Local Similarity	100.0%	Pred. No. 2.1e-139;		
Matches 456;	Conservative 0;	Mismatches 0;	Gaps 0.	

```

OY      1 ATGAAACCTTTAAAGGTGGACGACATTCGACGAATCGTAAGTTCTTGCGAGTCTTGAGCT 60
Db      1 ATGAAACCTTTAAAGGTGGACGACATTCGACGAATCGTAAGTTCTTGCGAGTCTTGAGCT 60
OY      61 GCGGCTCTATGATCAGCTGTTAACCCGTTGTTAACCCATGAATGGCAATGCATCCGGC 120
Db      61 GCGGCTCTATGATCAGCTGTTAACCCGTTGTTAACCCATGAATGGCAATGCATCCGGC 120
OY      121 CGGAGCTCAAGTTGAGAGATTATTCAGTACGGTTCGGTAAAGCTGCGCTTGCTCTGCAA 180
Db      121 CGGAGCTCAAGTTGAGAGATTATTCAGTACGGTTCGGTAAAGCTGCGCTTGCTCTGCAA 180
OY      181 AGCGATGCCCGCTAAATCTGAACACGACATTACCAGAGCGGTTATGGTAACGGCGCCGAT 240
Db      181 AGCGATGCCCGCTAAATCTGAACACGACATTACCAGAGCGGTTATGGTAACGGCGCCGAT 240
OY      241 GTAGGCCGAGGGTGGGGATATATGTACTATTGAACTGACTAGATGGTTTCAGAAATTAAT 300
Db      241 GTAGGCCGAGGGTGGGGATATATGTACTATTGAACTGACTAGATGGTTTCAGAAATTAAT 300
OY      301 GCCACCATTCGACCAAGTGGAAAGCGCTAAATACTCGGATTTATCTGTCGGCCCAATAGCGCGGT 360
Db      301 GCCACCATTCGACCAAGTGGAAAGCGCTAAATACTCGGATTTATCTGTCGGCCCAATAGCGCGGT 360
OY      361 AATTAACCCCGCGCTGCTTAATTCAGACCGCATCTGATTCAGAGCTAAATGTCGCTCAGGTT 420
Db      361 AATTAACCCCGCGCTGCTTAATTCAGACCGCATCTGATTCAGAGCTAAATGTCGCTCAGGTT 420
OY      421 GGTTTGGCAACAACGCCACGCGTAAACCAAGTATTA 456
Db      421 GGTTTGGCAACAACGCCACGCGTAAACCAAGTATTA 456

RESULT 2
/US-08-233-642A-56
/Sequence 56, Application US/08233642A
/GENERAL INFORMATION:
/  APPLICANT: Kay, William W.
/  APPLICANT: Collinson, S. Karen
/  APPLICANT: Cloutier, Sharon C.
/  TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
/  TITLE OF INVENTION: BASED VACCINES
/  NUMBER OF SEQUENCES: 58
/  NUMBER OF SEQUENCES: -
/  CORRESPONDENCE ADDRESS:
/  ADDRESSEE: Seed and Berry
/  STREET: 6300 Columbia Center, 701 Fifth Avenue
/  CITY: Seattle
/  STATE: Washington
/  COUNTRY: U.S.A.
/  ZIP: 98104-7092
/  COMPUTER READABLE FORM:
/  MEDIUM TYPE: Floppy disk
/  COMPUTER: IBM PC compatible
/  OPERATING SYSTEM: PC-DOS/MS-DOS
/  SOFTWARE: Patent In Release #1.0, Version #1.25
/  CURRENT APPLICATION DATA:
/  APPLICATION NUMBER: US/08/233,642A
/  FILING DATE: 26-APR-1994
/  CLASSIFICATION: 424
/  ATTORNEY/AGENT INFORMATION:
/  NAME: King, Joshua

```

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 88.4%; Score 403.2; DB 6; Length 456;
Best Local Similarity 92.8%; Pred. No. 6.1e-122;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
D 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCGGTGTTTACCAGTAATGSCATGATCCGGC 120
D 61 GCGCTATGATCAGCTGTTACCGGTGTTTACCAGTAATGSCATGATCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTACGCTGCTCTGCA 180
D 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
D 181 AGCGATGCCGTTAAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
D 241 GTAGGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTATGTCGCGCAATACGCGCGT 360
D 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTATGTCGCGCAATACGCGCGT 360
QY 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGTCGTCAGGTT 420
D 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAAGCCGAGGCTAACGATATTAA 456
D 421 GGTTTGGCAACAAGCCGAGGCTAACGATATTAA 456
```

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match 88.4%; Score 403.2; DB 23; Length 456;
Best Local Similarity 92.8%; Pred. No. 6.1e-122;
Matches 423; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
D 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCGGTGTTTACCAGTAATGSCATGATCCGGC 120
D 61 GCGCTATGATCAGCTGTTACCGGTGTTTACCAGTAATGSCATGATCCGGC 120
QY 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTACGCTGCTCTGCA 180
D 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTCCGTTACGCTGCTCTGCA 180
QY 181 AGCGATGCCGTTAAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
D 181 AGCGATGCCGTTAAATCTGAAACGACCTTACCCAGAGCGTTATGTAACGCGCGAT 240
QY 241 GTAGGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
D 241 GTAGGCCAGGGTGGGATTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTATGTCGCGCAATACGCGCGT 360
D 301 GCCACATCGACCAAGTGAACGCTAAACCTCCGATATTATGTCGCGCAATACGCGCGT 360
QY 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGTCGTCAGGTT 420
D 361 AATAACGCGCGCTGTTAATCAAGCCGATCTGATTCAGCGTATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAAGCCGAGGCTAACGATATTAA 456
D 421 GGTTTGGCAACAAGCCGAGGCTAACGATATTAA 456
```

RESULT 4
US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p73 from GP63 of Leishmania major.
US-09-543-407-19

Query Match 78.9%; Score 360; DB 23; Length 456;
Best Local Similarity 86.8%; Pred. No. 1.2e-107;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
D 1 ATGAACTTTTAAAGTGGCAGCATTCGCAATCGTAGTTTCTGCGAGTCTTGCT 60
QY 61 GCGCTATGATCAGCTGTTACCGGTGTTTACCAGTAATGSCATGATCCGGC 120
```

```
Db 61 GGGGTGTTCCACATGCGGGCGCGGCTATCATTAACGGCGCGCAATGTTCCGGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAAGCAGCATTTACCCAGAGGGTTATGTAACGGCGCGAT 240
Db 181 AGCGATGCCCGTAAATATATATATCAGTGTGTACCCGTTGTTACCATGAAATGCAAT 240
Qy 241 GTAGGCGCAGGTGCGGATATAGTACTATTTGAACTGACTCAGATGTTGTTGAAATAT 300
Db 241 GTAGGCGCAGGTGCGGATATAGTACTATTTGAACTGACTCAGATGTTGTTGAAATAT 300
Qy 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
Db 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
Qy 421 GGTTTGGCAACACGCGCAAGCTTAA 456
Db 421 GGTTTGGCAACACGCGCAAGCTTAA 456
```

RESULT 5

```
US-09-543-407-17
; Sequence 17, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-17
```

```
Query Match 77.5%; Score 353.6; DB 23; Length 456;
Best Local Similarity 86.0%; Pred. No. 1.6e-105;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
```

```
Qy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGCAATCTAGTTTCTGCGAGTGTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGCAATCTAGTTTCTGCGAGTGTGCTGCT 60
Qy 61 GGGCTCTATGATCAGCTGTGTTACCCGTTGTTATCCATGAAATGGCACATGTCATCCGG 120
Db 61 GGGCTCTCTCCACATGAGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCGGG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAAGCAGCATTTACCCAGAGGGTTATGTAACGGCGCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAAGCAGCATTTACCCAGAGGGTTATGTAACGGCGCGAT 240
```

```
Qy 241 GTAGGCGCAGGTGCGGATATAGTACTATTTGAACTGACTCAGATGTTTCAAGAAATAT 300
Db 241 GTAGGCGCAGGTGCGGATATATATGATCTATGAACTGACTCAGAAATGTTTCAAGAAATAT 300
Qy 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
Db 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
Qy 421 GGTTTGGCAACACGCGCAAGCTTAA 456
Db 421 GGTTTGGCAACACGCGCAAGCTTAA 456
```

RESULT 6

```
US-09-543-407-11
; Sequence 11, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-11
```

```
Query Match 77.2%; Score 352; DB 23; Length 456;
Best Local Similarity 85.7%; Pred. No. 5.3e-105;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
```

```
Qy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGCAATCTAGTTTCTGCGAGTGTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGCAATCTAGTTTCTGCGAGTGTGCTGCT 60
Qy 61 GGGCTCTATGATCAGCTGTGTTACCCGTTGTTATCCATGAAATGGCACATGTCATCCGG 120
Db 61 GGGCTCTCTCCACATGAGGGGCGGCGGTATCATTAACGGCGGCAATAGTTCGGG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATATCAGTACGTTCCGCTAACGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAAGCAGCATTTACCCAGAGGGTTATGTAACGGCGCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAAGCAGCATTTACCCAGAGGGTTATGTAACGGCGCGAT 240
Qy 241 GTAGGCGCAGGTGCGGATATAGTACTATTTGAACTGACTCAGATGTTTCAAGAAATAT 300
Db 241 GTAGGCGCAGGTGCGGATATATATGATCTATGAACTGACTCAGAAATGTTTCAAGAAATAT 300
Qy 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Db 301 GCGACCATGACACGAGTGAACGCTTAAATCTCCGATATTAATGTGCGCCAAATCGCGGT 360
Qy 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
Db 361 AATAAGCGCGCGTGGTTATATCAGACCGCATCTGATTCAGCGGTAAATGTCGTCAGTT 420
```

Qy 421 GGTTCGCAACAAGCGGCTTACCAAGATTAA 456
Db 421 GCACATGCAACAAGCGGCTTACCAAGATTAA 456

RESULT 7
US-09-543-407-13

; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-13

Query Match 77.2%; Score 352; DB 23; Length 456;

Best Local Similarity 85.7%; Pred. No. 5.3e-105; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 65;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCTATGATCAGCTGTTACCCGCTGTTTACCCATGAATGCGACATGCTCGGC 120
Db 61 GGCCTCTGTTCCACATGGGGCGGCGGCGGTATCATACGCGGCGCAATAGTTCCGCGC 120
Qy 121 CCGACCTCAAGCTTGAAGATTATCAGTACGCTTCCGTAACGCTGCTGCTGCAA 180
Db 121 CCGACCTCAAGCTTGAAGATTATCAGTACGCTTCCGTAACGCTGCTGCTGCAA 180
Qy 181 AGCGATGCCGCTTAATCTGAAGACCATTAACGAGCGGTTATGTTAAGCGGCGCAT 240
Db 181 AGCGATGCCGCTTAATCTGAAGACCATTAACGAGCGGTTATGTTAAGCGGCGCAT 240
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTCAGAAATGTTTCAAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTCAGAAATGTTTCAAAATAT 300
Qy 301 GCCACCATCGACGAGTGAAGCTAAATCTCCGATATTAATGTCGCGCATTAAGCGCGT 360
Db 301 GCCACCATCGACGAGTGAAGCTAAATCTCCGATATTAATGTCGCGCATTAAGCGCGT 360
Qy 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGCGCAT 420
Db 361 CTGGTACCCGCTGTTTATACCAATGAATGCGACATGCAACGTAATGTCGCGCAT 420
Qy 421 GGTTCGCAACAAGCGGCTTACCAAGATTAA 456
Db 421 GGTTCGCAACAAGCGGCTTACCAAGATTAA 456

RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of *Leishmania major*.
US-09-543-407-23

Query Match 77.2%; Score 352; DB 23; Length 456;

Best Local Similarity 85.7%; Pred. No. 5.3e-105; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 65;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTCTGCGAGTCTGCT 60
Qy 61 GGCCTCTATGATCAGCTGTTACCCGCTGTTTACCCATGAATGCGACATGCTCGGC 120
Db 61 GGCCTCTGTTCCACATGGGGCGGCGGCGGTATCATACGCGGCGCAATAGTTCCGCGC 120
Qy 121 CCGACCTCAAGCTTGAAGATTATCAGTACGCTTCCGTAACGCTGCTGCTGCAA 180
Db 121 CCGACCTCAAGCTTGAAGATTATCAGTACGCTTCCGTAACGCTGCTGCTGCAA 180
Qy 181 AGCGATGCCGCTTAATCTGAAGACCATTAACGAGCGGTTATGTTAAGCGGCGCAT 240
Db 181 AGCGATGCCGCTTAATCTGAAGACCATTAACGAGCGGTTATGTTAAGCGGCGCAT 240
Qy 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTCAGAAATGTTTCAAAATAT 300
Db 241 GTAGCCAGGGTGGGATTAATAGTACTTAACTGACTCAGAAATGTTTCAAAATAT 300
Qy 301 GCCACCATCGACGAGTGAAGCTAAATCTCCGATATTAATGTCGCGCATTAAGCGCGT 360
Db 301 GCCACCATCGACGAGTGAAGCTAAATCTCCGATATTAATGTCGCGCATTAAGCGCGT 360
Qy 361 AATAAGCGCGGCTGTTAATCAGACCGCATCTGATTCAGCGTAAATGTCGCGCAT 420
Db 361 CATGAATGCGACATGGAATCAGACCGCATCTGATTCAGCGTAAATGTCGCGCAT 420
Qy 421 GGTTCGCAACAAGCGGCTTACCAAGATTAA 456
Db 421 GGTTCGCAACAAGCGGCTTACCAAGATTAA 456

RESULT 9
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-25

Query Match
Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
QY 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
DB 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
QY 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTACCCAGAGCGGTATGTAAGCGCGCGAT 240
DB 181 CTGGTACCCGTTGTTACCATGAAATGGCAGCATGCTGTTATGTAAGCGCGCGAT 240
QY 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATGACGACGTAAGCAACGCTTAAACCTCCGATTAATGTAATGTAATGTAAT 360
DB 301 GCCACATGACGACGTAAGCAACGCTTAAACCTCCGATTAATGTAATGTAATGTAAT 360
QY 361 AATAAGCGCGCTGTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
DB 361 AATAAGCGCGCTGTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
QY 421 GGTGTTGGCAACGACGCGCTTAAACGCTTAAACGCTTAAACGCTTAAACGCTTAAAC 456
DB 421 GGTGTTGGCAACGACGCGCTTAAACGCTTAAACGCTTAAACGCTTAAACGCTTAAAC 456

RESULT 10
US-09-543-407-21

Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 76.5%; Score 348.8; DB 23; Length 456;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
QY 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
DB 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
QY 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180
QY 181 AGCGATGCCGTAATCTGAAACGACATTACCCAGAGCGGTATGTAAGCGCGCGAT 240
DB 181 AGCGATGCCGTAATCTGAAACGACATTACCCAGAGCGGTATGTAAGCGCGCGAT 240
QY 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATGACGACGTAAGCAACGCTTAAACCTCCGATTAATGTAATGTAATGTAAT 360
DB 301 GCCACATGACGACGTAAGCAACGCTTAAACCTCCGATTAATGTAATGTAATGTAAT 360
QY 361 AATAAGCGCGCTGTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
DB 361 AATAAGCGCGCTGTTATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 420
QY 421 GGTGTTGGCAACGACGCGCTTAAACGCTTAAACGCTTAAACGCTTAAACGCTTAAAC 456
DB 421 GGTGTTGGCAACGACGCGCTTAAACGCTTAAACGCTTAAACGCTTAAACGCTTAAAC 456

RESULT 11
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match
Best Local Similarity 76.5%; Score 348.8; DB 23; Length 456;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGATTGCGAGCAATGCTGTTCTGGCAGTCTGCTGCT 60
QY 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
DB 61 GGGCTCTATGATCAGCTGCTTACCGTGTGTTACCATGAAATGGCAGCATGCTGCTGCT 120
QY 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAACGTTGAGCATTATTCAGTACGTTCCGCTGTAACGCTGCTGCTGCTGCA 180

```

Db      121 CCGGACTCAAGCTTGAGCATTTATCATGTAACGGTTCGGCTAACGGCGCTTGCTTGCAA 180
Qy      181 AGCGATGCGCGCTTAATCTGAAACGACATTCACAGAGCGGTATGTGAACGGCGCGCAT 240
Db      181 AGCGATGCGCGCTTAATCTGAAACGACATTCACAGAGCGGTATGTGAACGGCGCGCAT 240
Qy      241 GTAGCGCGCGGTGCGGATTAATGACTTATGAACTGACTCAAGTAAGTGTGGAATAAT 300
Db      241 TATATCATGCTGCTGTTACCCGCTGTTGTTACCAATGAAATGCGACATGCAATGAATAAT 300
Qy      301 GCCACATCGACGACGTGGAACGCTAAAACCTCCGATATTAATGTCGCGCAATACGCGCGT 360
Db      301 GCCACATCGACGACGTGGAACGCTAAAACCTCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy      361 AATAACGCGCGCGCTGTTAATCAACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db      361 AATAACGCGCGCGCTGTTAATCAACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456
Db      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456

```

```

RESULT 12
US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PasteSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-29

```

```

Query Match      75.8%; Score 345.6; DB 23; Length 456;
Best Local Similarity 84.9%; Pred. No. 6.9e-103; Indels 0; Gaps 0;
Matches 387; Conservative 0; Mismatches 69;
Qy      1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTTCTGGAGTCTCTGGCT 60
Db      1 ATGAACTTTTAAAGTGGAGCATTCGACAAATCGAGTTTCTGGAGTCTCTGGCT 60
Qy      61 GCGCTATGATCACTGCTGTTAACCCGCTGTTTACCAATGAATGCGACATGCTCCGC 120
Db      61 GCGCTATGATCACTGCTGTTAACCCGCTGTTTACCAATGAATGCTCCGC 120
Qy      121 CCGGACTCAAGCTTGAGCATTTATCATGTAACGGTTCGGCTAACGGCGCTTGCTTGCAA 180
Db      121 CCGGACTCAAGCTTGAGCATTTATCATGTAACGGTTCGGCTAACGGCGCTTGCTTGCAA 180
Qy      181 AGCGATGCGCGCTTAATCTGAAACGACATTCACAGAGCGGTATGTGAACGGCGCGAT 240
Db      181 AGCGATGCGCGCTTAATCTGAAACGACATTCACAGAGCGGTATGTGAACGGCGCGAT 240
Qy      241 GTAGCGCGCGGTGCGGATTAATGACTTATGAACTGACTCAAGTAAGTGTGGAATAAT 300
Db      241 GTAGCGCGCGGTGCGGATTAATGACTTATGAACTGACTCAAGTAAGTGTGGAATAAT 300
Qy      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456
Db      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456

```

```

Qy      301 GCCACATCGACGACGTGGAACGCTAAAACCTCCGATATTAATGTCGCGCAATACGCGCGT 360
Db      301 GCCACATCGACGACGTGGAACGCTAAAACCTCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy      361 AATAACGCGCGCGCTGTTAATCAACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Db      361 AATAACGCGCGCGCTGTTAATCAACCGCATCTGATTCAGCGTAATGTCGTCAGGTT 420
Qy      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456
Db      421 GGTTCGCAACAAAGCGGCTAACCGATTTAA 456

```

```

RESULT 13
US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-Apr-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

```

```

Query Match      62.3%; Score 284.2; DB 6; Length 361;
Best Local Similarity 99.0%; Pred. No. 1.3e-82;
Matches 286; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy      110 ATGCATCGCGCGCGACTCAAGCTTGTGAGCATTTATCAGTACGGTTCGCTAACGCTGCGC 169
Db      47 ATAGTTCGCGCGCGCGACTCAAGCTTGTGAGCATTTATCAGTACGGTTCGCTAACGCTGCGC 106
Qy      170 TTGCTGCAAGAGGATGCCGTAATCTGAAACGACATTCACAGAGCGGTATGCTA 229
Db      107 TTGCTGCAAGAGGATGCCGTAATCTGAAACGACATTCACAGAGCGGTATGCTA 166

```

QY	230	ACGGCGCCGATNTATAGGCGACAGGCTGCCGATTAATAGTACTATTGACCTGACAGATGTT	283
Db	167	ACGGCGCCGATGTAGGCGACAGGCTGCCGATTAATAGTACTATTGACCTGACAGATGTT	226
QY	290	TCAGAAATTAATGCGACCATCGACAACTGGAACGCTTAAAACTCCGATTTACTGTGGCC	349
Db	227	TCAGAAATTAATGCGACCATCGACAACTGGAACGCTTAAAACTCCGATTTACTGTGGCC	286
QY	350	AATAACGGCGGTAATAACGCCGCGCTGGTAAATCAGACCGCATTCGATTC	398
Db	287	AATAACGGCGGTAATAACGCCGCGCTGGTAAATCAGACCGCATTCGATTC	335

RESULT 14
US-09-543-407-3

```

1 Sequence 3, Application US/09543407
2
3 GENERAL INFORMATION:
4
5 APPLICANT: White, Aaron P.
6 APPLICANT: Dorn, James L.
7 APPLICANT: Collinson, S. Karen
8 APPLICANT: Kay, William W.
9
10 TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
11 FILE REFERENCE: 920043.406
12
13 CURRENT APPLICATION NUMBER: US/09/543.407
14 CURRENT FILING DATE: 2000-04-05
15
16 NUMBER OF SEQ ID NOS: 59
17
18 SOFTWARE: FastSeq for Windows Version 4.0
19
20 SEQ ID NO 3
21
22 LENGTH: 456
23
24 TYPE: DNA
25
26 ORGANISM: E. COLI
27
28 US-09-543-407-3

```

Query Match	48.8%;	Score 222.4;	DB 23;	Length 456;
Best Local Similarity	68.0%;	Pred. No. 3.9e-62;		
Matches 310; Conservative	0;	Mismatches 146;	Indels 0;	Gaps 0

```

; Sequence 172, Application US/09252691B
; GENERAL INFORMATION:
; APPLICANT: Keith G. Weinstock et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROBACTERES
; TITLE OF INVENTION: CLOACAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.135
; CURRENT APPLICATION NUMBER: US/09/252,691B
; CURRENT FILING DATE: 1999-02-18
; NUMBER OF SEQ ID NOS: 11324
; SEQ ID NO 172
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Enterobacter cloacae
US-09-252-691-172

```

Query Match	48.8%;	Score 222.4;	DB 17;	Length 477;
Best Local Similarity	68.0%;	Pred. No. 4e-62;		
Matches 310; Conservative	0;	Mismatches 146;	Indels 0;	Gaps 0;

```
Search completed: March 18, 2004, 02:47:01
Job time : 2924.24 secs
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds

(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-15

Perfect score: 456
Sequence: 1 atgaacttttaaaagtgc.....ccacgctcaaccagctatca 456Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 200000000Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending Patents NA New: *
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq: *
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq: *
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq: *
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq: *
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq: *
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq: *
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	222.4	48.8	675	US-09-741-873C-3	Sequence 3, Appli
2	180	39.5	5	US-09-741-873C-1	Sequence 1, Appli
3	33.4	7.3	3990	US-09-830-230A-629	Sequence 629, App
4	32.4	7.1	150223	US-10-624-19A-1	Sequence 27, Appli
5	30.8	6.8	1575	US-10-788-792-27	Sequence 1, Appli
6	29.6	6.5	24900	US-10-767-471-10736	Sequence 10736, A
7	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, Ap
8	29.4	6.4	1400	US-60-545-213-6406	Sequence 6406, Ap
9	29.4	6.4	1400	US-60-545-213-6407	Sequence 6407, Ap
10	29.2	6.4	249	US-10-771-241-117	Sequence 117, App
11	29.2	6.4	757	US-10-771-241-29	Sequence 29, Appli
12	28.8	6.3	83009	US-10-417-375A-143	Sequence 143, App
13	28.8	6.3	2511	PCT-US04-05654-2123	Sequence 123, Ap
14	28.4	6.2	20043	US-10-021-688A-869	Sequence 869, App
15	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, App
16	28.2	6.2	201	US-60-548-091-22688	Sequence 22688, A
17	28.2	6.2	1295	US-10-100-683-5480	Sequence 5480, Ap
18	28.2	6.2	3416	US-10-100-683-3397	Sequence 3397, Ap
19	28.2	6.2	8206	US-10-021-688A-3529	Sequence 3529, Ap
20	28.2	6.2	2007	US-10-775-972-153	Sequence 153, App
21	28	6.1	2148	US-10-775-972-154	Sequence 154, App
22	27.8	6.1	639	US-10-767-701-7904	Sequence 7904, Ap
23	27.8	6.1	1422	US-10-767-795-7335	Sequence 7335, Ap
24	27.8	6.1	2342	US-10-451-467A-277	Sequence 277, App
25	27.8	6.1	2895	US-10-767-701-14458	Sequence 14458, A

27	27.8	6.1	18802	6	US-10-767-471-10574	Sequence 10574, A
28	27.8	6.1	32372	6	US-10-767-471-10611	Sequence 10611, A
29	27.8	6.1	1126118	6	US-10-767-471-10615	Sequence 10615, A
30	27.6	6.1	201	6	US-10-767-471-28293	Sequence 28293, A
31	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
32	27.6	6.1	600	7	US-60-545-213-6292	Sequence 6292, Ap
33	27.6	6.1	1498	6	US-10-100-683-5108	Sequence 5108, Ap
34	27.6	6.1	1523	6	US-10-100-683-750	Sequence 750, App
35	27.6	6.1	1662	6	US-10-100-683-5106	Sequence 5106, App
36	27.6	6.1	2571	5	US-09-999-183A-4	Sequence 4, Appli
37	27.6	6.1	3035	6	US-10-786-892-279	Sequence 279, App
38	27.6	6.1	232882	6	US-10-767-471-110679	Sequence 110679, A
39	27.4	6.0	201	6	US-10-767-471-11899	Sequence 11899, A
40	27.4	6.0	868	6	US-10-767-701-167	Sequence 167, App
41	27.4	6.0	318752	6	US-10-767-471-10553	Sequence 10553, A
42	27.2	6.0	959	6	US-10-767-701-1446	Sequence 1446, Ap
43	27.2	6.0	998	5	US-09-461-537A-3	Sequence 3, Appli
44	27	5.9	511	6	US-10-767-701-18763	Sequence 18763, A
45	27	5.9	987	6	US-10-100-683-3090	Sequence 3090, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Oleen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 48.8%; Score 222.4; DB 5; Length 675;
Best Local Similarity 68.0%; Pred. No. 8.7e-67;
Matches 310; Conservative 0; Mismatches 146; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGACGATTCGACGATTCGATTCGACGATTCGCTGCTGCT 60
DB 83 ATGAACCTTTAAAGTGGACGATTCGACGATTCGATTCGCTGCTGCTGCTGCTGCA 142
QY 61 GCGCTCATGATGCGTGTGTTACCCGCTGTTTACCATGAAATGCAATGATCCGGCC 120
DB 143 GGTGTTCCTTCGATGCGGCGGCGGCGGATACCAAGCGTGGCGGTAAATAGCGGC 202
QY 121 CCGACTCAACGTTGACGATTTACGATGCTTCGCTCAACGCTGCGCTGCTGCTGCA 180

```

Db      203 CCAATTCGAGCTGAACATTACAGTACGGTGGCGGTAATCTGCACTTGCTCGCA 262
Qy      181 AGCGATGCCCGTAATCTGAACACCATTAACCGAGCGGTAATGTAACGCCCGCAT 240
Db      263 ACTATGCCCCGTAATCTGACTTACTATTAACCGAGATGGCGGGTAATGTTGCAAT 322
Qy      241 GTAGCCGAGGTTGGGTAATTAATGACTATTAATGAATCTGAGAAATGTTGAGAATAT 300
Db      323 GTTGCTAGAGGCTGAGTATGACGTACATGATCTTACCCAAAGTGGCTTGGAACAGC 382
Qy      301 GCCACCATTCAGACGTGGAACGCTAAATCTCCGATATTAATCTGTCGGCAATAGCGCGT 360
Db      383 GCTACTCTTATCATGTGGAACGCGCAAAATTCGAATGACGTTAAACAGTTGCGTGT 442
Qy      361 AATAACGCCCGCTGTTATATCAGACCGCATCTGATTCAGGTAATGTTGCTGAGATT 420
Db      443 GGCAACGGTCTGCTCACTTACCAAGTGCATCTAACTCTCCGTCACAGTACAGATT 502
Qy      421 GGTTCGCAACACGCCACGCGCTAACAGTATTAA 456
Db      503 GGCTTTGTGAACACGCGACCGCTCATGACTATA 538

```

RESULT 2

```

; Sequence 1, Application US/09741873C
; GENERAL INFORMATION:
; APPLICANT: Normark, Staffan
; APPLICANT: Olsen, Arne
; TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
; FILE REFERENCE: 012889-084
; CURRENT APPLICATION NUMBER: US/09/741,873C
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: SE 8801723-1
; PRIOR FILING DATE: 1998-05-06
; PRIOR APPLICATION NUMBER: US 08/978,878
; PRIOR FILING DATE: 1997-11-26
; PRIOR APPLICATION NUMBER: US 07/347,189
; PRIOR FILING DATE: 1989-05-04
; PRIOR APPLICATION NUMBER: US 07/789,437
; PRIOR FILING DATE: 1991-11-06
; PRIOR APPLICATION NUMBER: US 07/970,846
; PRIOR FILING DATE: 1992-11-03
; PRIOR APPLICATION NUMBER: US 08/187,865
; PRIOR FILING DATE: 1994-01-28
; PRIOR APPLICATION NUMBER: US 08/318,519
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 396
; TYPE: DNA
; ORGANISM: Escherichia coli
US-09-741-873C-1

```

Query Match 39.5%; Score 180; DB 5; Length 396;
 Best Local Similarity 70.6%; Pred. No. 2.7e-52;
 Matches 240; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

Qy      117 CGGCCGAGCACTGTTAGCATTTATCGATGACGTTCCGCTAAGCGTGGCTGCTCT 176
Db      57 CGGCCCAATTCGATGAGCAATTTACAGTACGTTGCGGTAATCTGCACTTGCTCT 116
Qy      177 GCAAGGATGCGCGTAAATCTGAAGAGCAATTAACCGAGCGGTTATGTTAAAGCGGC 236
Db      117 GCAAAAGATGCGCGTAACTCTGACTTGAATTAACCGAGATGGCGGTTAATGGTGC 176
Qy      237 CGATGAGCGGACGAGTGGCGGATTAATGATCTATTGACTGACTGAGATGGTTGAGAAA 236
Db      177 AGATGTTGGTGAAGGCTCAGATGACAGCTCATGATCTGACCCAGCTGCTTCGTTAA 236
Qy      297 TAATGCCACATTCAGACGTGGAACGCTAAATCTCCGATTTTCTGTGGCCAAATACG 356

```

```

Db      237 CAGCGTACTCTTGTATCATGTGAACGCCAAATTTGAATGACGTTAAACAGTTCCG 296
Qy      357 CGGTATTAACGCCGCGCTGTTAATCAGACCGCATCTGATTCAGGTTAATGTTGCTCA 416
Db      297 TGTGTGCAACGGTGTGCGAGTTGACAGATGATCTCTCCGTCAACGTTACTCA 356
Qy      417 GGTGTTTGGCAACACGCCACGCTTAACAGTATTAA 456
Db      357 GGTGCTTTGTGAACACGACCGCTCATGACTATA 396

```

RESULT 3

```

; Sequence 629, Application US/09830230A
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Lyme Disease Vaccines
; FILE REFERENCE: PB481US
; CURRENT APPLICATION NUMBER: US/09/830,230A
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: PCT/US98/12718
; PRIOR FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: 60/057,483
; PRIOR FILING DATE: 1997-09-03
; PRIOR APPLICATION NUMBER: 60/053,344
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/053,377
; PRIOR FILING DATE: 1997-07-22
; PRIOR APPLICATION NUMBER: 60/050,359
; PRIOR FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 756
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 629
; LENGTH: 3990
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1143)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1148)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1210)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1244)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1247)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1250)
; OTHER INFORMATION: n equals a,t,g, or c
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1251)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-830-230A-629

```

Query Match

Best Local Similarity 7.3%; Score 33.4; DB 5; Length 3990;
 Matches 100; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

US-60-545-213-2134/c
; Sequence 2134, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2134
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2134

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCCCGCATGTAGCGCGGTGCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGCTTGCAGTACCGGTGACCGGAGTGGGATCGGATTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCGACCATGCGACGATGGAAGCTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCAGACCATCTCTCCACACTCTGGAAC 318
QY 332 CCATATTAATCTGTGCGGCAATAGCGCGGTATTAACCCCGGTGTTATCAGACCGCAT 391
DB 317 CCGTTGAGCAGAGCGGCGTCTGATGAGCATTAACCTCCCTGACTGCGATCAGGTTTCT 258
QY 392 CTGATTCACGCGTATGTCGTCAGGTTGTTTGGCAACAACCGCAAGGTTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTCGAAAGTGAATGAAGAGGGGTGACCTG 199

RESULT 8
US-60-545-213-2135/c
; Sequence 2135, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2135
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-2135

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCGCGCATGTAGCGCGGTGCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGCTTGCAGTACCGGTGACCGGAGTGGGATCGGATTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCGACCATGCGACGATGGAAGCTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCAGACCATCTCTCCACACTCTGGAAC 318
QY 332 CCATATTAATCTGTGCGGCAATAGCGCGGTATTAACCCCGGTGTTATCAGACCGCAT 391
DB 317 CCGTTGAGCAGAGCGGCGTCTGATGAGCATTAACCTCCCTGACTGCGATCAGGTTTCT 258

QY 392 CTGATTCACGCGTATGTCGCTAGGTTGTTTGGCAACAACCGCAGGCTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTCGAAAGTGAATGAAGAGGGGTGACCTG 199

RESULT 9
US-60-545-213-6406/c
; Sequence 6406, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6406
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6406

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCGCGCATGTAGCGCGGTGCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGCTTGCAGTACCGGTGACCGGAGTGGGATCGGATTATTA 378
QY 272 AACTGACTCAGATGGTTTCAGAAATATGCGACCATGCGACGATGGAAGCTAAAACT 331
DB 377 TTGGGGTTCCTCCATCCACTGCTTGAAGAGCAGACCATCTCTCCACACTCTGGAAC 318
QY 332 CCATATTAATCTGTGCGGCAATAGCGCGGTATTAACCCCGGTGTTATCAGACCGCAT 391
DB 317 CCGTTGAGCAGAGCGGCGTCTGATGAGCATTAACCTCCCTGACTGCGATCAGGTTTCT 258
QY 392 CTGATTCACGCGTATGTCGTCAGGTTGTTTGGCAACAACCGCAAGGTTAACCG 450
DB 257 GGGATTCCTCCAGTGTGTGATTAATTCGAAAGTGAATGAAGAGGGGTGACCTG 199

RESULT 10
US-60-545-213-6407/c
; Sequence 6407, Application US/60545213
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; APPLICANT: Mounts, William Martin
; TITLE OF INVENTION: Nucleic Acid Arrays for Monitoring Expression Profiles of Drug
; FILE REFERENCE: AM101083 (031896-042099)
; CURRENT APPLICATION NUMBER: US/60/545,213
; NUMBER OF SEQ ID NOS: 303284
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6407
; LENGTH: 1400
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-545-213-6407

Query Match
Best Local Similarity 6.4%; Score 29.4; DB 7; Length 1400;
Matches 108; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 212 CCCAGAGCGGTTATGTTAGCGCGCGCATGTAGCGCGGTGCGGATTAATAGTACTATTG 271
DB 437 CCTGGGCGATAGCGCTTGCAGTACCGGTGACCGGAGTGGGATCGGATTATTA 378

Page 5

APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc J
APPLICANT: Reuber, T. Lynne
APPLICANT: Keddie, James
APPLICANT: Dubell III, Arnold N
APPLICANT: Pineda, Omlira
APPLICANT: Repetti, Peter
APPLICANT: Guterson, Karen
APPLICANT: Yu, Guo-Liang
APPLICANT: Brown, Pierre E
APPLICANT: Kumamoto, Roderick W
APPLICANT: Pilgrim, Marsha L
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
FILE REFERENCE: MB1-0047 PCT
CURRENT APPLICATION NUMBER: PCT/US04/05654
PRIOR FILING DATE: 2004-03-03
PRIOR APPLICATION NUMBER: 10/374,780
PRIOR FILING DATE: 2003-02-25
PRIOR APPLICATION NUMBER: 10/675,852
NUMBER OF SEQ ID NOS: 2950
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2123
LENGTH: 2511
TYPE: DNA
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G1548 Predicted polypeptide sequence is paralogous to G390, G391
PCT-US04-05654-2123

Query Match

Best Local Similarity 6.2%; Score 28.4; DB 1; Length 2511;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 21 AGCATTGCGAGCATGTAAGTTCTGCGAGTCTGCTGCGCTGCTGATGATCAGCTGT 80
DB 1470 AGCAGCAGCAGTAAGTAGAGGCTTGTAGTCCCGAGTGGAGATTGGAGGGCAGGT 1529
QY 81 TACCCGTGTTGTTACCCATGAATGCGACATGCA 114
DB 1530 TAACTTCACCTGCTCATCATATGAGCATGAA 1563

RESULT 15

US-10-021-698A-869/C
Sequence 869, Application US/10021698A
GENERAL INFORMATION:

APPLICANT: KEITH, TIM
APPLICANT: LITTLE, RANDALL
APPLICANT: VAN EERDEWEGH, PAUL
APPLICANT: DUPUIS, JOSEE
APPLICANT: DEL MASTRO, RICHARD
APPLICANT: SIMON, JASON
APPLICANT: ALLEN, KRISTINA
TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
FILE REFERENCE: 2976-4044US1
CURRENT APPLICATION NUMBER: US/10/021,698A
PRIOR FILING DATE: 2001-10-22
PRIOR APPLICATION NUMBER: 60/211,749
NUMBER OF SEQ ID NOS: 6160
SOFTWARE: PatentIn 2.1
SEQ ID NO 869
LENGTH: 20043

TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (6103)
OTHER INFORMATION: a, t, c or g

FEATURE:
NAME/KEY: modified_base
LOCATION: (20008)
OTHER INFORMATION: a, t, c or g
US-10-021-698A-869

Query Match

Best Local Similarity 6.2%; Score 28.4; DB 6; Length 20043;
Matches 53; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 9 TTAAAGTGGCAGCATTCGACGATCGTAAGTTCTGCGAGTCTGCTGCGCTCTTA 68
DB 10361 TCTAATCTGTAATCTGCGCTGCTATGTTGGCTCCGCTGTCACAGCTGGGGCTG 10302
QY 69 TGATCAGCTGGTTACCCGCTGTTACCCATGAA 102
DB 10301 GGAAGAGAGTGTGCTGCTGTGGAATGCTGGGAA 10268

Search completed: March 17, 2004, 08:25:46
Job time : 71.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaactttaaagtcg.....ccagcgctacacagatrraa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Pending Patents NA Main:

1: /cgn2_6/ptodata/2/pna/US06_PCTUS_COMB.seq.*
2: /cgn2_6/ptodata/2/pna/US06_COMB.seq.*
3: /cgn2_6/ptodata/2/pna/US07_COMB.seq.*
4: /cgn2_6/ptodata/2/pna/US08_COMB.seq.*
5: /cgn2_6/ptodata/2/pna/US081_COMB.seq.*
6: /cgn2_6/ptodata/2/pna/US082_COMB.seq.*
7: /cgn2_6/ptodata/2/pna/US083_COMB.seq.*
8: /cgn2_6/ptodata/2/pna/US084_COMB.seq.*
9: /cgn2_6/ptodata/2/pna/US085_COMB.seq.*
10: /cgn2_6/ptodata/2/pna/US086_COMB.seq.*
11: /cgn2_6/ptodata/2/pna/US087_COMB.seq.*
12: /cgn2_6/ptodata/2/pna/US088_COMB.seq.*
13: /cgn2_6/ptodata/2/pna/US089_COMB.seq.*
14: /cgn2_6/ptodata/2/pna/US090_COMB.seq.*
15: /cgn2_6/ptodata/2/pna/US091_COMB.seq.*
16: /cgn2_6/ptodata/2/pna/US092_COMB.seq.*
17: /cgn2_6/ptodata/2/pna/US093_COMB.seq.*
18: /cgn2_6/ptodata/2/pna/US093A_COMB.seq.*
19: /cgn2_6/ptodata/2/pna/US093B_COMB.seq.*
20: /cgn2_6/ptodata/2/pna/US094_COMB.seq.*
21: /cgn2_6/ptodata/2/pna/US095A_COMB.seq.*
22: /cgn2_6/ptodata/2/pna/US095B_COMB.seq.*
23: /cgn2_6/ptodata/2/pna/US095C_COMB.seq.*
24: /cgn2_6/ptodata/2/pna/US095D_COMB.seq.*
25: /cgn2_6/ptodata/2/pna/US095E_COMB.seq.*
26: /cgn2_6/ptodata/2/pna/US096A_COMB.seq.*
27: /cgn2_6/ptodata/2/pna/US096B_COMB.seq.*
28: /cgn2_6/ptodata/2/pna/US096C_COMB.seq.*
29: /cgn2_6/ptodata/2/pna/US096D_COMB.seq.*
30: /cgn2_6/ptodata/2/pna/US096E_COMB.seq.*
31: /cgn2_6/ptodata/2/pna/US097A_COMB.seq.*
32: /cgn2_6/ptodata/2/pna/US097B_COMB.seq.*
33: /cgn2_6/ptodata/2/pna/US097C_COMB.seq.*
34: /cgn2_6/ptodata/2/pna/US098A_COMB.seq.*
35: /cgn2_6/ptodata/2/pna/US098B_COMB.seq.*
36: /cgn2_6/ptodata/2/pna/US098C_COMB.seq.*
37: /cgn2_6/ptodata/2/pna/US098D_COMB.seq.*
38: /cgn2_6/ptodata/2/pna/US099A_COMB.seq.*
39: /cgn2_6/ptodata/2/pna/US099B_COMB.seq.*
40: /cgn2_6/ptodata/2/pna/US099C_COMB.seq.*
41: /cgn2_6/ptodata/2/pna/US099E_COMB.seq.*
42: /cgn2_6/ptodata/2/pna/US099F_COMB.seq.*
43: /cgn2_6/ptodata/2/pna/US100A_COMB.seq.*

44: /cgn2_6/ptodata/2/pna/US100B_COMB.seq.*
45: /cgn2_6/ptodata/2/pna/US101A_COMB.seq.*
46: /cgn2_6/ptodata/2/pna/US101B_COMB.seq.*
47: /cgn2_6/ptodata/2/pna/US102A_COMB.seq.*
48: /cgn2_6/ptodata/2/pna/US102B_COMB.seq.*
49: /cgn2_6/ptodata/2/pna/US103A_COMB.seq.*
50: /cgn2_6/ptodata/2/pna/US103B_COMB.seq.*
51: /cgn2_6/ptodata/2/pna/US104A_COMB.seq.*
52: /cgn2_6/ptodata/2/pna/US104B_COMB.seq.*
53: /cgn2_6/ptodata/2/pna/US106_COMB.seq.*
54: /cgn2_6/ptodata/2/pna/US107A_COMB.seq.*
55: /cgn2_6/ptodata/2/pna/US107B_COMB.seq.*
56: /cgn2_6/ptodata/2/pna/US6000_COMB.seq.*
57: /cgn2_6/ptodata/2/pna/US6001_COMB.seq.*
58: /cgn2_6/ptodata/2/pna/US6002_COMB.seq.*
59: /cgn2_6/ptodata/2/pna/US6003_COMB.seq.*
60: /cgn2_6/ptodata/2/pna/US6004_COMB.seq.*
61: /cgn2_6/ptodata/2/pna/US6005_COMB.seq.*
62: /cgn2_6/ptodata/2/pna/US6006_COMB.seq.*
63: /cgn2_6/ptodata/2/pna/US6007_COMB.seq.*
64: /cgn2_6/ptodata/2/pna/US6008_COMB.seq.*
65: /cgn2_6/ptodata/2/pna/US6009_COMB.seq.*
66: /cgn2_6/ptodata/2/pna/US6010_COMB.seq.*
67: /cgn2_6/ptodata/2/pna/US6011_COMB.seq.*
68: /cgn2_6/ptodata/2/pna/US6012_COMB.seq.*
69: /cgn2_6/ptodata/2/pna/US6013_COMB.seq.*
70: /cgn2_6/ptodata/2/pna/US6014_COMB.seq.*
71: /cgn2_6/ptodata/2/pna/US6015_COMB.seq.*
72: /cgn2_6/ptodata/2/pna/US6016_COMB.seq.*
73: /cgn2_6/ptodata/2/pna/US6017_COMB.seq.*
74: /cgn2_6/ptodata/2/pna/US6018_COMB.seq.*
75: /cgn2_6/ptodata/2/pna/US6019_COMB.seq.*
76: /cgn2_6/ptodata/2/pna/US6020_COMB.seq.*
77: /cgn2_6/ptodata/2/pna/US6021_COMB.seq.*
78: /cgn2_6/ptodata/2/pna/US6022_COMB.seq.*
79: /cgn2_6/ptodata/2/pna/US6023A_COMB.seq.*
80: /cgn2_6/ptodata/2/pna/US6023B_COMB.seq.*
81: /cgn2_6/ptodata/2/pna/US6024_COMB.seq.*
82: /cgn2_6/ptodata/2/pna/US6025_COMB.seq.*
83: /cgn2_6/ptodata/2/pna/US6026_COMB.seq.*
84: /cgn2_6/ptodata/2/pna/US6027_COMB.seq.*
85: /cgn2_6/ptodata/2/pna/US6028_COMB.seq.*
86: /cgn2_6/ptodata/2/pna/US6029_COMB.seq.*
87: /cgn2_6/ptodata/2/pna/US6030_COMB.seq.*
88: /cgn2_6/ptodata/2/pna/US6031_COMB.seq.*
89: /cgn2_6/ptodata/2/pna/US6032_COMB.seq.*
90: /cgn2_6/ptodata/2/pna/US6033_COMB.seq.*
91: /cgn2_6/ptodata/2/pna/US6034_COMB.seq.*
92: /cgn2_6/ptodata/2/pna/US6035_COMB.seq.*
93: /cgn2_6/ptodata/2/pna/US6036_COMB.seq.*
94: /cgn2_6/ptodata/2/pna/US6037_COMB.seq.*
95: /cgn2_6/ptodata/2/pna/US6038_COMB.seq.*
96: /cgn2_6/ptodata/2/pna/US6039_COMB.seq.*
97: /cgn2_6/ptodata/2/pna/US6040_COMB.seq.*
98: /cgn2_6/ptodata/2/pna/US6041_COMB.seq.*
99: /cgn2_6/ptodata/2/pna/US6042_COMB.seq.*
100: /cgn2_6/ptodata/2/pna/US6043_COMB.seq.*
101: /cgn2_6/ptodata/2/pna/US6044_COMB.seq.*
102: /cgn2_6/ptodata/2/pna/US6045_COMB.seq.*
103: /cgn2_6/ptodata/2/pna/US6046_COMB.seq.*
104: /cgn2_6/ptodata/2/pna/US6047_COMB.seq.*
105: /cgn2_6/ptodata/2/pna/US6048_COMB.seq.*
106: /cgn2_6/ptodata/2/pna/US6049_COMB.seq.*
107: /cgn2_6/ptodata/2/pna/US6050_COMB.seq.*
108: /cgn2_6/ptodata/2/pna/US6051_COMB.seq.*
109: /cgn2_6/ptodata/2/pna/US6052_COMB.seq.*
110: /cgn2_6/ptodata/2/pna/US6053_COMB.seq.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23 US-09-543-407-17	Sequence 17, Appl
2	406.4	89.1	456	6 US-08-233-642A-56	Sequence 56, Appl
3	406.4	89.1	456	23 US-09-543-407-1	Sequence 1, Appl
4	363.2	79.6	456	23 US-09-543-407-19	Sequence 19, Appl
5	355.2	77.9	456	23 US-09-543-407-11	Sequence 11, Appl
6	355.2	77.9	456	23 US-09-543-407-13	Sequence 13, Appl
7	355.2	77.9	456	23 US-09-543-407-23	Sequence 23, Appl
8	353.6	77.5	456	23 US-09-543-407-25	Sequence 25, Appl
9	353.6	77.5	456	23 US-09-543-407-15	Sequence 15, Appl
10	352	77.2	456	23 US-09-543-407-21	Sequence 21, Appl
11	348.8	76.5	456	23 US-09-543-407-27	Sequence 27, Appl
12	282.2	61.9	361	6 US-08-233-642A-54	Sequence 54, Appl
13	230.4	50.5	456	23 US-09-543-407-3	Sequence 3, Appl
14	230.4	50.5	456	23 US-09-543-407-3	Sequence 3, Appl
15	230.4	50.5	456	23 US-09-543-407-3	Sequence 3, Appl
16	230.4	50.5	456	23 US-09-543-407-3	Sequence 3, Appl
17	228	50.0	453	92 US-60-352-946-1	Sequence 1, Appl
18	228	50.0	453	92 US-60-352-946-1	Sequence 1, Appl
19	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
20	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
21	219.2	48.1	477	17 US-09-252-691C-172	Sequence 172, App
22	186.4	40.9	396	13 US-08-978-878-1	Sequence 1, Appl
23	161.6	35.4	360	31 US-09-741-873B-1	Sequence 1, Appl
24	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
25	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
26	64.8	14.2	522	17 US-09-252-691C-170	Sequence 170, App
27	51.2	11.2	100	45 US-10-417-886-170	Sequence 170, App
28	50	11.0	78	23 US-09-543-407-52	Sequence 52, Appl
29	50	11.0	78	23 US-09-543-407-53	Sequence 53, Appl
30	48	10.5	48	23 US-09-543-407-9	Sequence 9, Appl
31	42.4	9.3	100	45 US-10-146-492B-78	Sequence 78, Appl
32	41.2	9.0	78	23 US-09-543-407-49	Sequence 49, Appl
33	40.8	8.9	100	45 US-10-146-492B-79	Sequence 79, Appl
34	40.6	8.9	100	45 US-10-146-492B-79	Sequence 79, Appl
35	39.4	8.6	4762	62 US-60-068-118-239	Sequence 239, App
36	39.4	8.6	4762	62 US-60-068-118-239	Sequence 239, App
37	39.4	8.6	2242716	33 US-09-806-866A-13	Sequence 13, Appl
38	39.4	8.6	2242716	33 US-09-806-866A-13	Sequence 13, Appl
39	39.2	8.6	2272325	43 US-10-018-470A-1	Sequence 1, Appl
40	37	8.1	14682	33 US-09-806-866A-30	Sequence 30, Appl
41	37	8.1	2242716	33 US-09-806-866A-30	Sequence 1068, App
42	37	8.1	2272325	43 US-10-018-470A-1	Sequence 1, Appl
43	36.4	8.0	561	17 US-09-252-691C-171	Sequence 171, App
44	36.4	8.0	561	17 US-09-252-691C-171	Sequence 171, App
45	36.4	8.0	561	17 US-10-417-886-171	Sequence 171, App

ALIGNMENTS

RESULT 1
US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Dorian, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
FILE REFERENCE: 920043.406
PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
CURRENT APPLICATION NUMBER: US/09/543,407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afaA
OTHER INFORMATION: Sequence containing the replacement fragment
US-09-543-407-17

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2.2e-133;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAACTTTTAAAGTGGAGCATTCGAGCATGTGTTCTGCGAGTCTGCT	60
DB	1	ATGAACTTTTAAAGTGGAGCATTCGAGCATGTGTTCTGCGAGTCTGCT	60
QY	61	GGCGTCTTCCAAATGGCGCGCGGTATCATACGCGCGCAATAGTCCGGC	120
DB	61	GGCGTCTTCCAAATGGCGCGCGGTATCATACGCGCGCAATAGTCCGGC	120
QY	121	CCGACTATATATAGCTGTGTTACCCGTGTGTTACCAATGAGCATGCTCAA	180
DB	121	CCGACTATATATAGCTGTGTTACCCGTGTGTTACCAATGAGCATGCTCAA	180
QY	181	AGCGATGCCGTTAACTGTAACGACATTCACGAGCGGTATGTTAAGCGCGCAT	240
DB	181	AGCGATGCCGTTAACTGTAACGACATTCACGAGCGGTATGTTAAGCGCGCAT	240
QY	241	GTTAGGCGAGGCGGTATGTTAAGTGAATGTAAGTGAATGTTCAAAATTAAT	300
DB	241	GTTAGGCGAGGCGGTATGTTAAGTGAATGTAAGTGAATGTTCAAAATTAAT	300
QY	301	GCCACCATGACAGTGTGTAAGTGAATGTTAAGTGAATGTTCAAAATTAAT	360
DB	301	GCCACCATGACAGTGTGTAAGTGAATGTTAAGTGAATGTTCAAAATTAAT	360
QY	361	ATAAGCGCGCGGTATGTTAAGTGAATGTTAAGTGAATGTTCAAAATTAAT	420
DB	361	ATAAGCGCGCGGTATGTTAAGTGAATGTTAAGTGAATGTTCAAAATTAAT	420
QY	421	GGTTTGGCAACAGCGCATGTAACGATTTAA 456	
DB	421	GGTTTGGCAACAGCGCATGTAACGATTTAA 456	

RESULT 2
US-08-233-642A-56
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Dorian, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESS: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURES:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 89.1%; Score 406.4; DB 6; Length 456;
Best Local Similarity 93.2%; Pred. No. 1.2e-117;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
QY 61 GGCGTGTTCACAAATGGGGCGGGCGGTATCATTAAGGGGGGCAATAGTTCCGGC 120
DB 61 GGCGTGTTCACAAATGGGGCGGGCGGTATCATTAAGGGGGGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATCAGCTGTTACCCGCTGTTGTTACCCAGTAAGTGAATGGCAGTGCAGCA 180
DB 121 CCGGACTATGATCAGCTGTTACCCGCTGTTGTTACCCAGTAAGTGAATGGCAGTGCAGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTTACCGGCCGAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTTACCGGCCGAT 240
QY 241 GTAGGCGCAGGGTGGGATTAATGACTATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGCAGGGTGGGATTAATGACTATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGACGTGAACGCTTAAATCCGATATTAATGTCGGCCAAATACGGCGGT 360
DB 301 GCCACATCGACGACGTGAACGCTTAAATCCGATATTAATGTCGGCCAAATACGGCGGT 360
QY 361 AATAAGCCCGGCTGTTAATCAGACCGCATGTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAAGCCCGGCTGTTAATCAGACCGCATGTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAGCGCAGCGCTAACAGATTTAA 456
DB 421 GGTTTGGCAACAGCGCAGCGCTAACAGATTTAA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1
Query Match 89.1%; Score 406.4; DB 23; Length 456;
Best Local Similarity 93.2%; Pred. No. 1.2e-117;
Matches 425; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
QY 61 GGCGTGTTCACAAATGGGGCGGGCGGTATCATTAAGGGGGGCAATAGTTCCGGC 120
DB 61 GGCGTGTTCACAAATGGGGCGGGCGGTATCATTAAGGGGGGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATCAGCTGTTACCCGCTGTTGTTACCCAGTAAGTGAATGGCAGTGCAGCA 180
DB 121 CCGGACTATGATCAGCTGTTACCCGCTGTTGTTACCCAGTAAGTGAATGGCAGTGCAGCA 180
QY 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTTACCGGCCGAT 240
DB 181 AGCGATGCCGCTTAATCTGAAACGACATTCACCGAGCGGTTATGTTACCGGCCGAT 240
QY 241 GTAGGCGCAGGGTGGGATTAATGACTATGTAATGTAATGTAATGTAATGTAATGTAAT 300
DB 241 GTAGGCGCAGGGTGGGATTAATGACTATGTAATGTAATGTAATGTAATGTAATGTAAT 300
QY 301 GCCACATCGACGACGTGAACGCTTAAATCCGATATTAATGTCGGCCAAATACGGCGGT 360
DB 301 GCCACATCGACGACGTGAACGCTTAAATCCGATATTAATGTCGGCCAAATACGGCGGT 360
QY 361 AATAAGCCCGGCTGTTAATCAGACCGCATGTGATTCAGCGTAATGTCGTCAGGTT 420
DB 361 AATAAGCCCGGCTGTTAATCAGACCGCATGTGATTCAGCGTAATGTCGTCAGGTT 420
QY 421 GGTTTGGCAACAGCGCAGCGCTAACAGATTTAA 456
DB 421 GGTTTGGCAACAGCGCAGCGCTAACAGATTTAA 456

RESULT 4
US-09-543-407-19
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding p73 from Gp63 of Leishmania major.
US-09-543-407-19

Query Match 79.6%; Score 363.2; DB 23; Length 456;
Best Local Similarity 87.3%; Pred. No. 6e-104;
Matches 398; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAAATCGTATGTTCTGGCAGTCTCTGGCT 60
QY 61 GGCGTGTTCACAAATGGGGCGGGCGGTATCATTAAGGGGGGCAATAGTTCCGGC 120

US-09-543-407-17.rtfm

```

Db      61  GGGCTGTTCCACATGGCGGCGCGGTATATATACGGCGCATATGTTCCGGC
Qy      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Db      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Qy      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Db      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Qy      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT

```

```

RESULT 5
US-09-543-407-11
Sequence 11, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FINGERPRINT SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: US/09/543.407
SOFTWARE: SEQ ID NOS: 2000-04-05
SEQ ID NO: 11
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atga
OTHER INFORMATION: Sequence containing the replacement fragment
US-09-543-407-11
Query Match
Best Local Similarity 77.9%; Score 355.2; DB 23; Length 456;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

```

Db      1  ATGAACTTTTAAAGTGGAGCATTTCCACATGTTGTTGTTGTTGTTGTTGTTGTTGTT
Qy      61  GGGCTGTTCCACATGGCGGCGCGGTATATATACGGCGCATATGTTCCGGC
Db      61  GGGCTGTTCCACATGGCGGCGCGGTATATATACGGCGCATATGTTGTTGTTGTTGTTGTTGTTGTT
Qy      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Db      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Qy      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Db      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Qy      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT

```

```

Qy      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Db      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Qy      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT

```

```

RESULT 6
US-09-543-407-13
Sequence 13, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FINGERPRINT SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: US/09/543.407
SOFTWARE: SEQ ID NOS: 2000-04-05
SEQ ID NO: 13
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b atga
OTHER INFORMATION: Sequence containing the replacement fragment
US-09-543-407-13
Query Match
Best Local Similarity 77.9%; Score 355.2; DB 23; Length 456;
Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

```

```

Db      1  ATGAACTTTTAAAGTGGAGCATTTCCACATGTTGTTGTTGTTGTTGTTGTTGTT
Qy      61  GGGCTGTTCCACATGGCGGCGCGGTATATATACGGCGCATATGTTCCGGC
Db      61  GGGCTGTTCCACATGGCGGCGCGGTATATATACGGCGCATATGTTGTTGTTGTTGTTGTTGTTGTT
Qy      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Db      121  CCGGACTATGATCGCTGTTACCGGTGTTGTTACCAATGGAATGCAATGACATGCA
Qy      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      181  AGGATGTCCTGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      241  GTAGGCGGAGGTGGGATATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Db      301  GCCACCATGACAGTGGAGAGCTTAAATCTCCGATTTGTTGCGCAATATGATATAT
Qy      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      361  AATAAGCCCGCGCTGTTATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT
Qy      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT
Db      421  GGTTCGCAACAGCGCAAGCTTACCAATGTAATGTAATGTAATGTAATGTAATGTAAT

```

Qy 421 GATTGGCAACACGCGCTAACGATATTA 456
Db 421 GATTGGCAACACGCGCTAACGATATTA 456

RESULT 7
US-09-543-407-23

; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCGTAGTTCTGGAGTCTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCGTAGTTCTGGAGTCTCTGCT 60
Qy 61 GCGCTGTTCCACATGGGGGGGGGGGGTATCATTAAGGGGGGGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGGGGGGGGTATCATTAAGGGGGGGGCAATGTTCCGGC 120
Qy 121 CCGGACTATGATGACGCTGTTACCGGTGTTGCCATGAATGGGCAATGAGCTGGA 180
Db 121 CCGGACTATGATGACGCTGTTACCGGTGTTGCCATGAATGGGCAATGAGCTGGA 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAAGCGGCGCAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAAGCGGCGCAT 240
Qy 241 GTAGGCGGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
Db 241 GTAGGCGGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
Qy 301 GCCACCATGACGAGTGAAGCTAAATCTCCGATATTACTGTCGGGCAATAGCGCGT 360
Db 301 GCCACCATGACGAGTGAAGCTAAATCTCCGATATTACTGTCGGGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGTCGTTAATCAGACCGCATCTGATTCACGTAATGTCGTCAGGTT 420
Db 361 AATAAGCGCGGTCGTTAATCAGACCGCATCTGATTCACGTAATGTCGTCAGGTT 420
Qy 421 GATTGGCAACACGCGCTAACGATATTA 456
Db 421 GATTGGCAACACGCGCTAACGATATTA 456

RESULT 8
US-09-543-407-25
; Sequence 25, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 25
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-25

Query Match 77.9%; Score 355.2; DB 23; Length 456;
Best Local Similarity 86.2%; Pred. No. 2.1e-101;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCGTAGTTCTGGAGTCTCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGAGCATCGTAGTTCTGGAGTCTCTGCT 60
Qy 61 GCGCTGTTCCACATGGGGGGGGGGGGTATCATTAAGGGGGGGGCAATGTTCCGGC 120
Db 61 GCGCTGTTCCACATGGGGGGGGGGGGTATCATTAAGGGGGGGGCAATGTTCCGGC 120
Qy 121 CCGGACTATGATGACGCTGTTACCGGTGTTGCCATGAATGGGCAATGAGCTGGA 180
Db 121 CCGGACTATGATGACGCTGTTACCGGTGTTGCCATGAATGGGCAATGAGCTGGA 180
Qy 181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAAGCGGCGCAT 240
Db 181 AGCGATGCGCGTAAATCTGAAACGACATTCACGAGCGGTTATGTAAGCGGCGCAT 240
Qy 241 GTAGGCGGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
Db 241 GTAGGCGGAGGTCGGATTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 300
Qy 301 GCCACCATGACGAGTGAAGCTAAATCTCCGATATTACTGTCGGGCAATAGCGCGT 360
Db 301 GCCACCATGACGAGTGAAGCTAAATCTCCGATATTACTGTCGGGCAATAGCGCGT 360
Qy 361 AATAAGCGCGGTCGTTAATCAGACCGCATCTGATTCACGTAATGTCGTCAGGTT 420
Db 361 AATAAGCGCGGTCGTTAATCAGACCGCATCTGATTCACGTAATGTCGTCAGGTT 420
Qy 421 GATTGGCAACACGCGCTAACGATATTA 456
Db 421 GATTGGCAACACGCGCTAACGATATTA 456

RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR
; TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-15

Query Match
Best Local Similarity 77.5%; Score 353.6; DB 23; Length 456;
Matches 392; Conservative 0; Mismatches 64; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
DB 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
QY 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
DB 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 180
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 180
QY 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
DB 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
QY 301 GCCACCATGACCGAGTGAACGCTAAACCTCCGATATTAAGTGGGCAATACGGCGGT 360
DB 301 GCCACCATGACCGAGTGAACGCTAAACCTCCGATATTAAGTGGGCAATACGGCGGT 360
QY 361 AATAACCGCGCGGTGTTATCATACCGCATTCGATTCAGGTTATGTTGCTGAGTT 420
DB 361 AATAACCGCGCGGTGTTATCATACCGCATTCGATTCAGGTTATGTTGCTGAGTT 420
QY 421 GGTGTTGGCAACAGCGGCTAATCAACGATTTAA 456
DB 421 GGTGTTGGCAACAGCGGCTAATCAACGATTTAA 456
```

RESULT 10
US-09-543-407-21
Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
DB 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
QY 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
DB 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
QY 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 180
DB 181 AGCGATGCCGTTAAATCTGAAACGACATTAACCGAGCGGTTATGTTAAGCGCGCAT 180
QY 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
DB 241 GTAGGCGAGGGTGGGATATGTAATGTAATGTAATGTAATGTAATGTAATGTAAT 240
QY 301 GCCACCATGACCGAGTGAACGCTAAACCTCCGATATTAAGTGGGCAATACGGCGGT 360
DB 301 GCCACCATGACCGAGTGAACGCTAAACCTCCGATATTAAGTGGGCAATACGGCGGT 360
QY 361 AATAACCGCGCGGTGTTATCATACCGCATTCGATTCAGGTTATGTTGCTGAGTT 420
DB 361 AATAACCGCGCGGTGTTATCATACCGCATTCGATTCAGGTTATGTTGCTGAGTT 420
QY 421 GGTGTTGGCAACAGCGGCTAATCAACGATTTAA 456
DB 421 GGTGTTGGCAACAGCGGCTAATCAACGATTTAA 456
```

RESULT 11
US-09-543-407-27
Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match
Best Local Similarity 77.2%; Score 352; DB 23; Length 456;
Matches 391; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
QY 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
DB 1 ATGAACTTTTAAAGGAGCATTCGAGCATGTAAGTTCTGAGAGTCTGAGCT 60
QY 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
DB 61 GGGGTGCTTCCAAATGGGGGCGGCGGTATCATACGGGGGCAATAGTCCGGC 120
QY 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
DB 121 CCGGACTATGATCAGCTGGTTACCGGTGTGTACCATGAAATGGCAATGATCCGGC 120
```

Db 121 CCGGACTGAGGTTAGCATTTATCATGACGGTCCGCTAACGGCTCGCTCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTATACGGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTATACGGCCGAT 240
Qy 241 GTAGGCCAGGGGCGCGATTAATGACTATGACGACTCAGAAAGTTTCAGAAATAT 300
Db 241 TATGATGACGTGGTTACCCGCTGTGTATCCCTAGAAAGGACATTCAGAAATAT 300
Qy 301 GCCACCATGACGACGATGGAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCGAT 360
Db 301 GCCACCATGACGACGATGGAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCGAT 360
Qy 361 AATAAGCGCGGCTGTTAATCATGACCGCATCTGATTCAGAGGTAATGTGCTCAGGTT 420
Db 421 GATTGGCAACAACGCGCATACCATGATTAA 456
Qy 421 GATTGGCAACAACGCGCATACCATGATTAA 456
Db 421 GATTGGCAACAACGCGCATACCATGATTAA 456

RESULT 12

US-09-543-407-29
Sequence 29, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT FILING DATE: 2000-04-05
CURRENT APPLICATION NUMBER: US/09/543.407
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-29

Query Match 76.5%; Score 348.8; DB 23; Length 456;
Best Local Similarity 85.3%; Pred. No. 2.2e-99;
Matches 389; Conservative 0; Mismatches 67; Indels 0; Gaps 0;
Qy 1 ATGAAACTTTTAAAGTGGAGCATTCGACGAAATCGATGTTCTGCGAGTCTGCT 60
Db 1 ATGAAACTTTTAAAGTGGAGCATTCGACGAAATCGATGTTCTGCGAGTCTGCT 60
Qy 61 GCGCTGTTCCAAATGCGGCGCGCGGCTAATCATTAACGCGCGGCAATAGTTCCGGC 120
Db 61 GCGCTGTTCCAAATGCGGCGCGCGGCTAATCATTAACGCGCGGCAATAGTTCCGGC 120
Qy 121 CCGGACTATGATCAAGCTGTTACCGGTTGTTACCCATGAAATGGCAATGCACTGCA 180
Db 121 CCGGACTATGATCAAGCTGTTATAGTACGTTCCGCTAACCGCTGCTGCTGCA 180
Qy 181 AGCGATGCCCGTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTATACGGCCGAT 240
Db 181 AGCGATGCCCGTAAATCTGAAAAGCACTTACCCAGAGCGGTTATGTATACGGCCGAT 240
Qy 241 GTAGGCCAGGGTGGCGATTAATGACTATGACGACTCAGAAAGTTTCAGAAATAT 300
Db 241 GTAGGCCAGGGTGGCGATTAATGACTATGACGACTCAGAAAGTTTCAGAAATAT 300

Qy 301 GCCACCATGACGACGATGGAACGCTTAAAACTCCGATATTATCTGCGCCAAATACGGCGAT 360
Db 301 GCCACCATGATGACGCTGTTACCCGCTGTGTATCCCTAGAAAGGACATGACGCGGAT 360
Qy 361 AATAAGCGCGGCTGTTAATCATGACCGCATCTGATTCAGAGGTAATGTGCTCAGGTT 420
Db 361 AATAAGCGCGGCTGTTAATCATGACCGCATCTGATTCAGAGGTAATGTGCTCAGGTT 420
Qy 421 GATTGGCAACAACGCGCATACCATGATTAA 456
Db 421 GATTGGCAACAACGCGCATACCATGATTAA 456

RESULT 13

US-08-233-642A-54
Sequence 54, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
APPLICANT: Doran, James L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR *SALMONELLA*-
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233.642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua
REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEEDANBERRY
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 361 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..357
US-08-233-642A-54

Query Match 61.9%; Score 282.2; DB 6; Length 361;
Best Local Similarity 90.1%; Pred. No. 2.8e-78;
Matches 302; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Qy 64 GTGCTTCCAAATGCGGCGCGGCTAATCATTAACGCGCGGCAATAGTTCCGGCCG 123
Db 1 GTGCTTCCAAATGCGGCGCGGCTAATCATTAACGCGCGGCAATAGTTCCGGCCG 60
Qy 124 GACTATGATCAAGCTGTTACCCGTTGTTACCATGAAATGGCAATGCACTGCAAGC 183
Db 61 GACTCAACGTTGACATTTATCATGACGTTCCGCTAACGCTGCTGCTCAAGC 120

```
QY 184 GATGCCCGTAATCTGAAACGACATTCACAGAGCGGTATGTGTAACGGCGCCGATGTA 243
DB 121 GATGCCCGTAATCTGAAACGACATTCACAGAGCGGTATGTGTAACGGCGCCGATGTA 180
QY 244 GGCCAGGCTGGGATATAGTACTATTAAGTACTGACATGAGTATGTTTCAAGAAATATGCC 303
DB 181 GGCCAGGCTGGGATATAGTACTATTAAGTACTGACATGAGTATGTTTCAAGAAATATGCC 240
QY 304 ACCATCCGACGATGAGACGCTTAAACCTCGATATTAATGCTGGCCCAATAGCGCGGTAAT 363
DB 241 ACCATCCGACGATGAGACGCTTAAACCTCGATATTAATGCTGGCCCAATAGCGCGGTAAT 300
QY 364 AACCCCGGCTGGTATATCAGACGCGCATCTGATTC 398
DB 301 AACCCCGGCTGGTATATCAGACGCGCATCTGATTC 335
```

```
RESULT 14
US-09-543-407-3
; Sequence 3, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543,407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: PatSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 456
; TYPE: DNA
; ORGANISM: E. coli
US-09-543-407-3
```

```
Query Match 50.5%; Score 230.4; DB 23; Length 456;
Best Local Similarity 69.1%; Pred. No. 8.2e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGGACGACATTCGACGAAATCGATGTTTCTGGCAGTCTTGAGCT 60
DB 1 ATGAACCTTTTAAAGTGGGACGACATTCGACGAAATCGATGTTTCTGGCAGTCTTGAGCT 60
QY 61 GGCGTCTTCCACAAATGGGGCGGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
DB 61 GGCGTCTTCCACAAATGGGGCGGCGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTATGATCGAGCTGTTACCCGTTGTTTACCCATGAAATGGACATGACCTGCA 180
DB 121 CCGGACTATGATCGAGCTGTTACCCGTTGTTTACCCATGAAATGGACATGACCTGCA 180
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
QY 241 ACTATGCCGCTAACTGACTGATTAATTAACGATGCGCGGTAATGTTGCAAGT 240
DB 241 ACTATGCCGCTAACTGACTGATTAATTAACGATGCGCGGTAATGTTGCAAGT 240
QY 241 GTAGCCAGGCTGGGATATAGTACTATTAATGAACTGACTGAGATGATGTTTCAAGAAATAT 300
DB 241 GTAGCCAGGCTGGGATATAGTACTATTAATGAACTGACTGAGATGATGTTTCAAGAAATAT 300
QY 301 GCCACATCGACGATGAGACGCTTAAACCTCGATATTAATGTTCCGCCCAATAGCGCGGT 360
DB 301 GCCACATCGACGATGAGACGCTTAAACCTCGATATTAATGTTCCGCCCAATAGCGCGGT 360
QY 361 GCTACTCTTATAGTGGAAACGCGCAAAATTTCTGAATATGACGTTTAAACAGTTCCGTGT 360
DB 361 GCTACTCTTATAGTGGAAACGCGCAAAATTTCTGAATATGACGTTTAAACAGTTCCGTGT 360
QY 361 AATTAAGCCCGCTGGTATATCAGACGCACTGATTCAGCGTATAGTGTGTCAGGTT 420
DB 361 AATTAAGCCCGCTGGTATATCAGACGCACTGATTCAGCGTATAGTGTGTCAGGTT 420
QY 421 GGTTTTGGCAACAAGCGCAAGGCTAACCGATATTA 456
DB 421 GGTTTTGGCAACAAGCGCAAGGCTAACCGATATTA 456
QY 421 GGCTTTGGTAAACAAGCGCAAGGCTAACCGATATTA 456
DB 421 GGCTTTGGTAAACAAGCGCAAGGCTAACCGATATTA 456
```

```
RESULT 15
US-08-978-878-3
; Sequence 3, Application US/08978878
; GENERAL INFORMATION:
; APPLICANT: NORMARK, Staffan
; APPLICANT: OLSEN, Arne
; TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
; FILE REFERENCE: 012889-081
; CURRENT APPLICATION NUMBER: US/08/978,878
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: SE 8801723-1
; EARLIER FILING DATE: 1988-05-06
; EARLIER APPLICATION NUMBER: US 07/347,189
; EARLIER FILING DATE: 1989-05-04
; EARLIER APPLICATION NUMBER: US 07/789,437
; EARLIER FILING DATE: 1991-11-06
; EARLIER APPLICATION NUMBER: US 07/970,846
; EARLIER FILING DATE: 1992-11-03
; EARLIER APPLICATION NUMBER: US 08/187,865
; EARLIER FILING DATE: 1994-01-28
; EARLIER APPLICATION NUMBER: US 08/318,519
; EARLIER FILING DATE: 1994-10-05
; EARLIER APPLICATION NUMBER: US 08/495,959
; EARLIER FILING DATE: 1995-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 675
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (83)..(538)
US-08-978-878-3
```

```
Query Match 50.5%; Score 230.4; DB 13; Length 675;
Best Local Similarity 69.1%; Pred. No. 9.4e-62;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

QY 1 ATGAACCTTTTAAAGTGGGACGACATTCGACGAAATCGATGTTTCTGGCAGTCTTGAGCT 60
DB 83 ATGAACCTTTTAAAGTGGGACGACATTCGACGAAATCGATGTTTCTGGCAGTCTTGAGCT 142
QY 61 GGCGTCTTCCACAAATGGGGCGGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 120
DB 143 GGCGTCTTCCACAAATGGGGCGGCGGTATCATATAAGCGCGGCAATAGTTCCGGC 202
QY 121 CCGGACTATGATCGAGCTGTTACCCGTTGTTTACCCATGAAATGGACATGACCTGCA 180
DB 203 CCGGACTATGATCGAGCTGTTACCCGTTGTTTACCCATGAAATGGACATGACCTGCA 262
QY 181 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 240
DB 263 AGCGATGCCGCTAAATCTGAAACGACCATTAACGAGCGGTTATGTTAAGCGCGCGAT 322
QY 241 GTAGCCAGGCTGGGATATAGTACTATTAATGAACTGACTGAGATGATGTTTCAAGAAATAT 300
DB 323 GTAGCCAGGCTGGGATATAGTACTATTAATGAACTGACTGAGATGATGTTTCAAGAAATAT 382
QY 301 GCCACATCGACGATGAGACGCTTAAACCTCGATATTAATGTTCCGCCCAATAGCGCGGT 360
DB 383 GCTACTCTTATAGTGGAAACGCGCAAAATTTCTGAATATGACGTTTAAACAGTTCCGTGT 442
QY 361 AATTAAGCCCGCTGGTATATCAGACGCACTGATTCAGCGTATAGTGTGTCAGGTT 420
DB 443 GCGAAGGCTGTCAGTATGACGACGATGATCTAATCTCTCGTCAAGTACTAGGTT 502
QY 421 GGTTTTGGCAACAAGCGCAAGGCTAACCGATATTA 456
DB 503 GGCTTTGGTAAACAAGCGCAAGGCTAACCGATATTA 538
```

Thu Mar 18 12:28:01 2004

us-09-543-407-17.rnp

Page 9

Search completed: March 18, 2004, 02:47:06
Job time : 2927.24 secs

THIS PAGE IS BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds

(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-17

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgc.....ccacgctaccagctatataa 456

Scoring table: IDENTITY_NUC

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Pending Patents NA New:*
1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	230.4	50.5	675	US-09-741-873C-3	Sequence 3, Appli
2	186.4	40.9	396	US-09-741-873C-1	Sequence 1, Appli
3	34.2	7.5	1355	US-10-045-674A-593	Sequence 593, App
4	31.4	6.9	435	US-10-021-658A-1216	Sequence 1216, Ap
5	31.4	6.9	2007	US-10-775-972-153	Sequence 153, App
6	31.4	6.9	2148	US-10-775-972-154	Sequence 154, App
7	31.2	6.8	9278	US-10-021-658A-796	Sequence 796, App
8	31	6.8	1099	US-10-767-701-8827	Sequence 8827, Ap
9	31	6.8	150223	US-10-624-149A-1	Sequence 1, Appli
10	30.8	6.8	1575	US-10-788-782-21	Sequence 27, Appli
11	30.4	6.7	983	US-10-767-701-1287	Sequence 1287, Ap
12	30.2	6.6	889	US-10-767-795-6329	Sequence 6329, Ap
13	30	6.6	1203	US-10-767-701-12486	Sequence 12486, A
14	29.8	6.5	1228	US-10-767-701-6233	Sequence 6233, Ap
15	29.6	6.5	656	US-10-767-701-20611	Sequence 20611, A
16	29.6	6.5	24900	US-10-767-471-10736	Sequence 10736, A
17	29.6	6.5	186449	US-10-021-658A-715	Sequence 715, App
18	29.4	6.4	300	US-10-767-701-28437	Sequence 28437, A
19	29.4	6.4	1400	US-10-767-701-17781	Sequence 17781, A
20	29.4	6.4	1400	US-60-545-213-2134	Sequence 2134, Ap
21	29.4	6.4	1400	US-60-545-213-2135	Sequence 2135, Ap
22	29.4	6.4	1400	US-60-545-213-6406	Sequence 6406, Ap
23	29.4	6.4	1400	US-60-545-213-6407	Sequence 6407, Ap
24	29.4	6.4	145597	US-10-624-149A-2	Sequence 2, Appli
25	29.4	6.4	145597	US-10-624-149A-2	Sequence 117, App
26	29.2	6.4	249	US-10-771-241-117	Sequence 117, App

27	29.2	6.4	757	US-10-771-241-29	Sequence 29, Appl
28	29	6.4	594	US-10-767-701-15041	Sequence 15041, A
29	29	6.4	1002	PCT-US04-05854-859	Sequence 859, App
30	29	6.4	1052	US-10-767-701-12607	Sequence 12607, A
31	29	6.4	2600	US-10-451-467A-39	Sequence 39, Appl
32	28.8	6.3	959	US-10-767-701-1446	Sequence 1446, Ap
33	28.8	6.3	83009	US-10-417-375A-143	Sequence 143, App
34	28.6	6.3	592	US-10-767-701-22979	Sequence 22979, A
35	28.6	6.3	749	US-10-767-701-12516	Sequence 12516, A
36	28.6	6.3	1186	PCT-US04-05654-1030	Sequence 1030, Ap
37	28.4	6.2	378	US-10-417-884A-1059	Sequence 1059, Ap
38	28.4	6.2	394468	US-60-548-091-5725	Sequence 5725, Ap
39	28.2	6.2	201	US-60-548-091-32268	Sequence 2268, A
40	28.2	6.2	610	US-10-767-701-8781	Sequence 8781, Ap
41	28.2	6.2	1295	US-10-100-683-5480	Sequence 5480, Ap
42	28.2	6.2	3416	US-10-100-683-3397	Sequence 3397, Ap
43	28.2	6.2	3776	US-60-548-091-34	Sequence 34, Appl
44	28.2	6.2	3828	US-60-548-091-38	Sequence 38, Appl
45	28.2	6.2	8206	US-10-021-658A-3529	Sequence 3529, Ap

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Oleen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 50.5%; Score 230.4; DB 5; Length 675;
Best Local Similarity 69.1%; Pred. No. 1,1e-68;
Matches 315; Conservative 0; Mismatches 141; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGAGCAATTCGAGTTCGAGTCTGAGCT 60
DB 83 ATGAACCTTTAAAGTGGAGCAATTCGAGTTCGAGTCTGAGCT 142
QY 61 GGCCTGTCACCAATGGCGGCGGCGGTAATTAACGCGGCGCAATGTTCCGCC 120
DB 143 GGTGTGTCTCTCGTACGGCGGCGGCGGTAACCAAGGTGTGCGGTAATTAAGCGGC 202
QY 121 CCGACTATGATGACGTGCTTACCCGCTGTTGTTACCATGAAATGACATGACATGCA 180

Db	181	GTGTGTCAGGGCTCATATGACAGTTCATTCGATCCCAACGTGGCTGGTAAAGC	240
QY	301	GCACCACTGACACAGTGGAAAGCTTAAATCTCCGATATTACTGTCCGACCAATACGGCGGT	360
Db	241	GCTACTCTTGTATCATAGGGAACGGCAAAAATTCGTAAATGACGGTGTAAACAGTTCCGGTGT	300
QY	361	AATTAACCGCGCGGTGGTTATTCAGACCGCATCTGATTCACGCGTAATGATGGCGTCAGGTT	420
Db	301	GGCAACGGTGTGACAGTTGACCAAGCTCATCTTAATCTCTCCGTCAACGTGACTCAGGTT	360
QY	421	GGTTTGGCAACACAGCCACGGCTTAACAGATTA	456
Db	361	GGCTTGGTAAACACGACCGCTCATCAGTACTAA	396
RESULT 3			
US-10-045-674A-593			
; Sequence 593, Application US/10045674A			
; GENERAL INFORMATION:			
; APPLICANT: LADNER, ROBERT C.			
; APPLICANT: COHEN, ROBERT U.			

RESULT 3
US-10-045-674A-593

```

? Sequence 593, Application US/10045674A
? GENERAL INFORMATION:
? APPLICANT: LADNER, ROBERT C.
? APPLICANT: COHEN, EDWARD H.
? APPLICANT: MASTRI, HORACIO G.
? APPLICANT: ROOKEY, KRISTIN L.
? APPLICANT: HOET, RENE
? APPLICANT: HOOGENBOOM, HENDRICUS R. J. M.
? TITLE OF INVENTION: NOVEL METHODS OF CONSTRUCTING LIBRARIES COMPRISING
? TITLE OF INVENTION: DISPLAYED AND/OR EXPRESSED MEMBERS OF A DIVERSE FAMILY
? TITLE OF INVENTION: OF PEPTIDES, POLYPEPTIDES OR PROTEINS AND THE NOVEL
? FILE REFERENCE: DYAK/002 CIP2
? CURRENT APPLICATION NUMBER: US/10/045,674A
? CURRENT FILING DATE: 2001-10-25
? PRIOR APPLICATION NUMBER: 06/198,069
? PRIOR FILING DATE: 2000-04-17
? PRIOR APPLICATION NUMBER: 09/837,306
? PRIOR FILING DATE: 2001-04-17
? NUMBER OF SEQ ID NOS: 635
? SOFTWARE: PatentIn Ver. 2.1
? SEQ ID NO 593
? LENGTH: 1355
? TYPE: DNA
? ORGANISM: Artificial Sequence
? FEATURE:
? OTHER INFORMATION: Description of Artificial Sequence: M13-III
? FEATURE:
? NAME/KEY: CDS
? LOCATION: (1)..(1305)
? US-10-045-674A-593

```

7.5%; Score 34.2; DB 6; Length 1355.

	1597	1607	1617	1627	1637	1647	1657	1667	1677	1687	1697	1707	1717	1727	1737	1747	1757	1767	1777	1787	1797	1807	1817	1827	1837	1847	1857	1867	1877	1887	1897	1907	1917	1927	1937	1947	1957	1967	1977	1987	1997	2007	2017	2027	2037	2047	2057	2067	2077	2087	2097	2107	2117	2127	2137	2147	2157	2167	2177	2187	2197	2207	2217	2227	2237	2247	2257	2267	2277	2287	2297	2307	2317	2327	2337	2347	2357	2367	2377	2387	2397	2407	2417	2427	2437	2447	2457	2467	2477	2487	2497	2507	2517	2527	2537	2547	2557	2567	2577	2587	2597	2607	2617	2627	2637	2647	2657	2667	2677	2687	2697	2707	2717	2727	2737	2747	2757	2767	2777	2787	2797	2807	2817	2827	2837	2847	2857	2867	2877	2887	2897	2907	2917	2927	2937	2947	2957	2967	2977	2987	2997	3007	3017	3027	3037	3047	3057	3067	3077	3087	3097	3107	3117	3127	3137	3147	3157	3167	3177	3187	3197	3207	3217	3227	3237	3247	3257	3267	3277	3287	3297	3307	3317	3327	3337	3347	3357	3367	3377	3387	3397	3407	3417	3427	3437	3447	3457	3467	3477	3487	3497	3507	3517	3527	3537	3547	3557	3567	3577	3587	3597	3607	3617	3627	3637	3647	3657	3667	3677	3687	3697	3707	3717	3727	3737	3747	3757	3767	3777	3787	3797	3807	3817	3827	3837	3847	3857	3867	3877	3887	3897	3907	3917	3927	3937	3947	3957	3967	3977	3987	3997	4007	4017	4027	4037	4047	4057	4067	4077	4087	4097	4107	4117	4127	4137	4147	4157	4167	4177	4187	4197	4207	4217	4227	4237	4247	4257	4267	4277	4287	4297	4307	4317	4327	4337	4347	4357	4367	4377	4387	4397	4407	4417	4427	4437	4447	4457	4467	4477	4487	4497	4507	4517	4527	4537	4547	4557	4567	4577	4587	4597	4607	4617	4627	4637	4647	4657	4667	4677	4687	4697	4707	4717	4727	4737	4747	4757	4767	4777	4787	4797	4807	4817	4827	4837	4847	4857	4867	4877	4887	4897	4907	4917	4927	4937	4947	4957	4967	4977	4987	4997	5007	5017	5027	5037	5047	5057	5067	5077	5087	5097	5107	5117	5127	5137	5147	5157	5167	5177	5187	5197	5207	5217	5227	5237	5247	5257	5267	5277	5287	5297	5307	5317	5327	5337	5347	5357	5367	5377	5387	5397	5407	5417	5427	5437	5447	5457	5467	5477	5487	5497	5507	5517	5527	5537	5547	5557	5567	5577	5587	5597	5607	5617	5627	5637	5647	5657	5667	5677	5687	5697	5707	5717	5727	5737	5747	5757	5767	5777	5787	5797	5807	5817	5827	5837	5847	5857	5867	5877	5887	5897	5907	5917	5927	5937	5947	5957	5967	5977	5987	5997	6007	6017	6027	6037	6047	6057	6067	6077	6087	6097	6107	6117	6
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

```

RESULT 5
US-10-775-972-153/C
: Sequence 153, Application US/10775972
: GENERAL INFORMATION:
: APPLICANT: Henderson, Robert A.
: APPLICANT: Wang, Tongtong
: APPLICANT: Bangur, Chaltanya S.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
: OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
: FILE REFERENCE: 210121.455621
: CURRENT APPLICATION NUMBER: US/10/775,972
: CURRENT FILING DATE: 2004-02-10
: NUMBER OF SEQ ID NOS: 563
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 153
: LENGTH: 2007
: TYPE: DNA
: ORGANISM: Homo sapiens
US-10-775-972-153

```

RESULT 7
 S-10-021-698A-796/c
 Sequence 796, Application US/10021698A
 GENERAL INFORMATION:
 APPLICANT: KEITH, TIM
 APPLICANT: LITTLE, RANDALL
 APPLICANT: VAN EERDEWEGH, PAUL
 APPLICANT: DUPUIS, JOSEF
 APPLICANT: DEL MASTRO, RICHARD
 APPLICANT: SIMON, JASON

[illegible]

```

; RESULT 8
; US-10-767-701-8827
; Sequence 8827, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 8827
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(1099)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28WAY03-CLUS59062_1
; US-10-767-701-8827

```

Query Match	6.8%;	Score 31;	DB 6;	Length 1099
Best Local Similarity	50.0%;	Pred. No. 1;		
Matches	76;	Conservative		

OY	99	CGCGCGCGGCATATAGTTCCGCGCCCGAGCTATGATCAGCTGGTATACCGTGTGTGTAACCA	158
Db	277	CGGCGGGGGGCTCGCGGCGGTTCTTGATCCGCGGGAGGAGAGCTACCTGGGTGGCGCCAA	336
OY	159	TGAAATGSCATGCACTGCGAAGGAGTGCCTGGAATCTGAAACGACATTAACCGAG	218

Db 337 TGCAGCCGAGACCCGAGCCGCGAGCGCCGACGATGCGAAGCGAAGCTGCCCTGCGCAGC 396
 Oy 219 CGGTTATGTTAAGCGGCGCGATGTAGGCCAAG 250
 Db 397 CGGTACGAGCTCGGCGCGCTGTCGTGCGGGCGG 428

RESULT 5
US-10-62

```

1 / Sequence 1, Application US/10624149A
2 / GENERAL INFORMATION:
3 / APPLICANT: Neubauer, Antonie
4 / APPLICANT: Ziegler, Christina
5 / TITLE OP INVENTION: gm-Negative EHV-Mutants without Heterologous Elements
6 / FILE REFERENCE: 1/1372
7 / CURRENT APPLICATION NUMBER: US/10/624,149A
8 / PRIOR FILING DATE: 2003-07-21
9 / PRIOR FILING DATE: 2002-08-14
10 / PRIOR FILING DATE: 2002-08-14
11 / PRIOR FILING DATE: 2002-07-19
12 / PRIOR APPLICATION NUMBER: DE 10233064
13 / PRIOR APPLICATION NUMBER: DE 10317008
14 / PRIOR FILING DATE: 2003-04-11
15 / NUMBER OF SEQ ID NOS: 2
16 / SOFTWARE: PatentIn Version 3.0
17 / SEQ ID NO 1
18 / LENGTH: 150223
19 / TYPE: DNA
20 / ORGANISM: Equine herpesvirus 1
21 / PUBLICATION INFORMATION:
22 / AUTHORS: Telford, E.A.
23 / AUTHORS: Watson, M.S.
24 / AUTHORS: McBride, K.
25 / AUTHORS: Davison, A.J.
26 / TITLE: The DNA sequence of equine herpesvirus-1
27 / JOURNAL: Virology
28 / VOLUME: 189
29 / ISSUE: 1
30 / PAGES: 304-316
31 / DATE: JUL-1992
32 / DATABASE ACCESSION NUMBER: NC 001491, NCBI
33 / DATABASE ENTRY DATE: 2000-08-01
34 / US-10-624-149A-1

```

Query Match	6.8%;	Score 31;	DB 6;	Length 150223;
Best Local Similarity	51.0%;	Pred. No. 11;		
Matches	73;	Conservative	0.0;	Matched

	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	534	535	536	537	538	539	540	541	542	543	544	545	546	547	548	549	550	551	552	553	554	555	556	557	558	559	560	561	562	563	564	565	566	567	568	569	570	571	572	573	574	575	576	577	578	579	580	581	582	583	584	585	586	587	588	589	590	591	592	593	594	595	596	597	598	599	600	601	602	603	604	605	606	607	608	609	610	611	612	613	614	615	616	617	618	619	620	621	622	623	624	625	626	627	628	
--	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	--

RESULT 10
US-10-788

```

: Sequence 27, Application US/10788792
: GENERAL INFORMATION:
: APPLICANT: Bayer Pharmaceuticals Corporation
: APPLICANT: Eweleigh, Deepa
: APPLICANT: Bigwood, Douglas
: TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
: FILE REFERENCE: 5152
: CURRENT APPLICATION NUMBER: US/10/788,792
: CURRENT FILING DATE: 2004-02-27
: PRIOR APPLICATION NUMBER: US 60/450,655

```

;; PRIOR FILING DATE: 2003-02-28
;; NUMBER OF SEQ ID NOS: 254
;; SOFTWARE: PatentIn version 3.2
;; SEQ ID NO: 27
;; LENGTH: 1575
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-10-767-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.5;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 317 GGAAGCTAAAGCTCCGATATTACTGCGGCATACGCGGTAAACGCGCGCTGG 376
Db 688 GGATTCATACAGCTCACTGACGCGTTCCTCACTTGGTCACAAAGACGTCGCGCTGG 629
Qy 377 TTAATCAGACCGCATCTGATTCAGCGTAAATGTCGCTCAGGTTGG 422
Db 628 TGCACCTGCTGGCTCTTCTCCAAAGACTTGGAGCTCAGGTTGG 583

RESULT 11
US-10-767-701-1287/C
; Sequence 1287, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 1287
; LENGTH: 883
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ..(883)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS19621_1
US-10-767-701-1287

Query Match 6.7%; Score 30.4; DB 6; Length 883;
Best Local Similarity 61.2%; Pred. No. 1.5;
Matches 49; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

Qy 52 GCTCTGCTGCGCTGCTTCCCAATGGGCGCGCGGTAAATCATACGCGCGGCAT 111
Db 493 GGTCTGATGATGACCTCATCTCTCAACGCTAGCGATACCGTTGTGCGCGAGT 434
Qy 112 AGTCCGCGCGCGACTATGA 131
Db 433 CTTTCGCGCGCTTGGCGATGA 414

RESULT 12
US-10-767-795-6329/C
; Sequence 6329, Application US/10767795
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53534)B
; CURRENT APPLICATION NUMBER: US/10/767,795
; CURRENT FILING DATE: 2004-01-30
; NUMBER OF SEQ ID NOS: 117596

;; SEQ ID NO 6329
;; LENGTH: 989
;; TYPE: DNA
;; ORGANISM: Gossypium hirsutum
;; FEATURE:
; OTHER INFORMATION: Clone ID: GOSHI-09MAV01-CL002_1
US-10-767-795-6329

Query Match 6.6%; Score 30.2; DB 6; Length 989;
Best Local Similarity 49.7%; Pred. No. 1.9;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

Qy 24 ATTGCAGCATCGATGTTCTGCGAGGCTCTGCGCGGTGCTGTCACAAATGGCGGG 83
Db 842 ATTAACGAAACAGACATCTTAGATTAATGATCTCTGCGCTTGTGCGCGATGATGA 783
Qy 84 CGGCGTAATCATACGCGCGCAATAGTTCGCGCGCGGACTATGATGATGCTGTAC 143
Db 782 TGTGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 723
Qy 144 CCGTGTGTATCCATGAAATGCAATGCACTGC 178
Db 722 CCGTATGAAACGACGAGCTGACCAACATAGC 688

RESULT 13
US-10-767-701-12486/C
; Sequence 12486, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 12486
; LENGTH: 1203
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) ..(1203)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAV03-CLUS13462_1
US-10-767-701-12486

Query Match 6.6%; Score 30; DB 6; Length 1203;
Best Local Similarity 48.3%; Pred. No. 2.4;
Matches 84; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

Qy 20 CAGATTGCGACATGATGTTCTGCGAGTCTGCGCGGTGCTGTCACAAATGGG 79
Db 271 CCGGACCGGCTTAAGTGCAGGTGCGCGCGGTGCTGTTGATGATGATGATGATG 212
Qy 80 CGGCGCGGTAATCATACGCGCGCAATAGTTCGCGCGCGGACTATGATGATGATG 139
Db 211 CCGGCGCGGAGTGCAGCGCGCGGTGTTGTGTGATGATGATGATGATGATGATG 152
Qy 140 TTACCGGTGTGTATCCATGAAATGCGACATGCACTGCAAGAGGATGCCGTA 193
Db 151 ATGAGCTGCAAGTGCAGCGCGCGCGCTGCGACCGGAGGAAGTGTGCTTGA 98

RESULT 14
US-10-767-701-6233/C
; Sequence 6233, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B

; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 6233
; LENGTH: 1228
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: SORBI-28MAY03-CLUS37536_1
US-10-767-701-6233

Query Match 6.5%; Score 29.8; DB 6; Length 1228;
Best Local Similarity 52.4%; Pred. No. 2.8;
Matches 64; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

QY 59 CTGGCCCTCTTCCACATATGGGCGCGCGGATATCATACGCGCGGCATATAGTTCG 118
DB 666 CTGCGCTGAGCGGAGATGAGCGGCGGATGATGTTCTTGCGGAGGCCCTCGGCGAG 607
QY 119 GCCCGACTATGATCAGCTGCTTACCCGTTGTTACCATGAAATGGACATGCACTGC 178
DB 606 GCGTCATCATGATGAGAGCTCATAGATTCTGTTGCTGAGGCGAGGTGGCTGC 547
QY 179 A 179
DB 546 A 546

RESULT 15
US-10-767-701-20611/C
; Sequence 20611, Application US/10767701
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof For Plant Improvement
; FILE REFERENCE: 38-21(5355)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 20611
; LENGTH: 656
; TYPE: DNA
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: 11064914
US-10-767-701-20611

Query Match 6.5%; Score 29.6; DB 6; Length 656;
Best Local Similarity 52.4%; Pred. No. 2.5;
Matches 65; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 28 GCAGCAATCGTAGTTCTGGCAGTCTCTGCTGCGCTGCGTTCACACATGGGCGGCGG 87
DB 126 GTAGAACTTGCGAGCCTGCGCTTGTGAGTCTGCGGCATCATGGGTGGCGTCC 67
QY 88 GGTATCATAAAGCGCGCGGCATAGTTCGCGCCGACTATGATCAGCTGTACCCGT 147
DB 66 GGTGACCTTGTCAGCGCGCGAGCGGAGCTGACGAGAACCCAGCGACAGCGT 7
QY 148 GTTG 151
DB 6 GCTG 3

Search completed: March 17, 2004, 08:25:47
Job time: 70.6301 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 17, 2004, 16:50:01 ; Search time 2921.94 Seconds
(without alignments)
5491.177 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttttaaaagtcgsc.....ccacgcttaaccagctatataa 456

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 37577330 seqs, 17593059518 residues

Total number of hits satisfying chosen parameters: 75154660

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

Pending Patente NA Main:*

- 1: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 2: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 3: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 4: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 5: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 6: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 7: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 8: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 9: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 10: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 11: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 12: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 13: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 14: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 15: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 16: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 17: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 18: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 19: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 20: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 21: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 22: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 23: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 24: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 25: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 26: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 27: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 28: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 29: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 30: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 31: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 32: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 33: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 34: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 35: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 36: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 37: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 38: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 39: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 40: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 41: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 42: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 43: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*

- 44: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 45: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 46: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 47: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 48: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 49: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 50: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 51: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 52: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 53: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 54: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 55: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 56: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 57: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 58: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 59: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 60: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 61: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 62: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 63: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 64: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 65: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 66: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 67: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 68: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 69: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 70: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 71: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 72: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 73: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 74: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 75: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 76: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 77: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 78: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 79: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 80: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 81: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 82: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 83: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 84: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 85: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 86: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 87: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 88: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 89: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 90: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 91: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 92: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 93: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 94: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 95: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 96: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 97: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 98: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 99: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 100: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 101: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 102: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 103: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 104: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 105: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 106: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 107: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 108: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 109: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*
- 110: /cgn2_6/prodata/2/pna/US0954_COMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	456	100.0	456	23	Sequence 19, Appl
2	412.8	90.5	456	6	Sequence 36, Appl
3	412.8	90.5	456	23	Sequence 1, Appl
4	385.6	84.6	456	23	Sequence 25, Appl
5	363.2	79.3	456	23	Sequence 17, Appl
6	361.6	79.3	456	23	Sequence 11, Appl
7	361.6	79.3	456	23	Sequence 13, Appl
8	361.6	79.3	456	23	Sequence 23, Appl
9	360	78.9	456	23	Sequence 15, Appl
10	360	78.9	456	23	Sequence 27, Appl
11	358.4	78.6	456	23	Sequence 21, Appl
12	355.2	77.9	456	23	Sequence 29, Appl
13	288.6	63.3	361	6	Sequence 54, Appl
14	236.8	51.9	456	23	Sequence 3, Appl
15	236.8	51.9	456	23	Sequence 3, Appl
16	236.8	51.9	456	23	Sequence 3, Appl
17	234.4	51.4	453	92	Sequence 1, Appl
18	234.4	51.4	453	92	Sequence 1, Appl
19	227.2	49.8	477	17	Sequence 172, Appl
20	227.2	49.8	477	17	Sequence 172, Appl
21	227.2	49.8	477	17	Sequence 172, Appl
22	192.8	42.3	336	13	Sequence 172, Appl
23	168	36.8	360	31	Sequence 1, Appl
24	64.8	14.2	522	17	Sequence 170, Appl
25	64.8	14.2	522	17	Sequence 170, Appl
26	64.8	14.2	522	51	Sequence 170, Appl
27	51.2	11.2	100	45	Sequence 77, Appl
28	50	11.0	78	23	Sequence 52, Appl
29	50	11.0	78	23	Sequence 52, Appl
30	48	10.5	78	23	Sequence 9, Appl
31	45	9.9	78	23	Sequence 9, Appl
32	43.4	9.5	78	23	Sequence 49, Appl
33	42.4	9.3	100	45	Sequence 48, Appl
34	42	9.2	2000	36	Sequence 5263, Appl
35	36.8	8.1	1290	37	Sequence 5066, Appl
36	36.8	8.1	1290	37	Sequence 5066, Appl
37	36.2	7.9	632	33	Sequence 23539, A
38	36.2	7.9	805	26	Sequence 61621, A
39	36.2	7.9	805	30	Sequence 61323, A
40	36.2	7.9	805	33	Sequence 61323, A
41	35.8	7.9	387	37	Sequence 21737, A
42	35.8	7.9	387	37	Sequence 21737, A
43	35.8	7.9	1281	20	Sequence 7264, A
44	35	7.6	542	24	Sequence 14113, A
45	34.8	7.6	542	24	Sequence 14113, A

ALIGNMENTS

RESULT 1
Sequence 19, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: PatSeq for Windows Version 4.0
SEQ ID NO 19
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-19

Query Match 100.0%; Score 456; DB 23; Length 456;
Best Local Similarity 100.0%; Pred. No. 2, 1e-136;
Matches 456; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 ATGAACTTTTAAAGTGGAGCATTCGACGATCGTGTTCGCGAGTGTCTGGCT 60
1 ATGAACTTTTAAAGTGGAGCATTCGACGATCGTGTTCGCGAGTGTCTGGCT 60
61 GGGCTCTTCCACATGCGCGCGCGCGGTATCATACGCGCGCGCAATAGTCCGCG 120
61 GGGCTCTTCCACATGCGCGCGCGCGGTATCATACGCGCGCGCAATAGTCCGCG 120
121 CCGAGCTCAACGTTGAGCATTTATCAGTACGCTGCTTACCGGTGTTACCGAATGACAT 180
121 CCGAGCTCAACGTTGAGCATTTATCAGTACGCTGCTTACCGGTGTTACCGAATGACAT 180
181 AGGATGCCCGTAAATATATATAGTACGCTGCTTACCGGTGTTACCGAATGACAT 240
181 AGGATGCCCGTAAATATATATAGTACGCTGCTTACCGGTGTTACCGAATGACAT 240
241 GCAGGCGGAGGCGGAT 300
241 GCAGGCGGAGGCGGAT 300
301 GCACCATGACGACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
301 GCACCATGACGACGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 360
361 AATAAGCGCGCGGTAT 420
361 AATAAGCGCGCGGTAT 420
421 GGTGTTGCAACAAAGCGGCGGCTTACGATTTAA 456
421 GGTGTTGCAACAAAGCGGCGGCTTACGATTTAA 456

RESULT 2
Sequence 56, Application US/08233642A
GENERAL INFORMATION:
APPLICANT: Kay, William W.
APPLICANT: Collinson, S. Karen
APPLICANT: Clouthier, Sharon C.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-
NUMBER OF SEQUENCES: 58
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: U.S.A.
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/233,642A
FILING DATE: 26-APR-1994
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: King, Joshua

REGISTRATION NUMBER: 35,570
REFERENCE/DOCKET NUMBER: 920043.403C3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
TELEX: 3723836 SEDANBERRY
INFORMATION FOR SEQ ID NO: 56:
SEQUENCE CHARACTERISTICS:
LENGTH: 456 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 1..456
US-08-233-642A-56

Query Match 90.5%; Score 412.8; DB 6; Length 456;
Best Local Similarity 94.1%; Pred. No. 2.1e-122;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGAATAGTCCGGC 120
DB 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATATGATCAGCTGGTACCGGTGTTGTTACCATGAAATGSCACAT 240
DB 181 AGCGATGCCCGTAAATATGATCAGCTGGTACCGGTGTTGTTACCATGAAATGSCACAT 240
QY 241 GCAGGCCAGGGGGGGGATATATGACTATTTGAATGACTGACCTGGAATGGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGGGGGGATATATGACTATTTGAATGACTGACCTGGAATGGTTTCAGAAATAT 300
QY 301 GCCACCATCGACCAAGTGAACGCTAAACCTCCATATTACTGTCGGCCAAATACGGCGGT 360
DB 301 GCCACCATCGACCAAGTGAACGCTAAACCTCCATATTACTGTCGGCCAAATACGGCGGT 360
QY 361 AATAACGCCGGCGCTGTTATATCAAGCCGATCTGATTCGACGTAATGCTGCTGCA 420
DB 361 AATAACGCCGGCGCTGTTATATCAAGCCGATCTGATTCGACGTAATGCTGCTGCA 420
QY 421 GGTTTGGCAACAACGCCACGGCTAACCAATTTAA 456
DB 421 GGTTTGGCAACAACGCCACGGCTAACCAATTTAA 456

RESULT 3
US-09-543-407-1
Sequence 1, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 456
TYPE: DNA
ORGANISM: Salmonella enteritidis

US-09-543-407-1

Query Match 90.5%; Score 412.8; DB 23; Length 456;
Best Local Similarity 94.1%; Pred. No. 2.1e-122;
Matches 429; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGAATAGTCCGGC 120
DB 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGAATAGTCCGGC 120
QY 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGCTTGAGCATTTATCATGACGTTCCGCTAACGCTGCTGCTGCA 180
QY 181 AGCGATGCCCGTAAATATGATCAGCTGGTACCGGTGTTGTTACCATGAAATGSCACAT 240
DB 181 AGCGATGCCCGTAAATATGATCAGCTGGTACCGGTGTTGTTACCATGAAATGSCACAT 240
QY 241 GCAGGCCAGGGGGGGGATATATGACTATTTGAATGACTGACCTGGAATGGTTTCAGAAATAT 300
DB 241 GTAGGCCAGGGGGGGGATATATGACTATTTGAATGACTGACCTGGAATGGTTTCAGAAATAT 300
QY 301 GCCACCATCGACCAAGTGAACGCTAAACCTCCATATTACTGTCGGCCAAATACGGCGGT 360
DB 301 GCCACCATCGACCAAGTGAACGCTAAACCTCCATATTACTGTCGGCCAAATACGGCGGT 360
QY 361 AATAACGCCGGCGCTGTTATATCAAGCCGATCTGATTCGACGTAATGCTGCTGCA 420
DB 361 AATAACGCCGGCGCTGTTATATCAAGCCGATCTGATTCGACGTAATGCTGCTGCA 420
QY 421 GGTTTGGCAACAACGCCACGGCTAACCAATTTAA 456
DB 421 GGTTTGGCAACAACGCCACGGCTAACCAATTTAA 456

RESULT 4
US-09-543-407-25
Sequence 25, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
TITLE OF INVENTION: PRESENTATION OF HETEROLOGOUS PEPTIDE SEQUENCES
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 25
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding P73 from GP63 of Leishmania major.
US-09-543-407-25

Query Match 84.6%; Score 385.6; DB 23; Length 456;
Best Local Similarity 90.4%; Pred. No. 1.4e-113;
Matches 412; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

QY 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
DB 1 ATGAACTTTTAAAGTGGCAGCATTCGACGAATCGTATGTTCTGGCAGTCTCTGCT 60
QY 61 GGGCGTGTTCACAAATGGGGGGGGGGGGTATCATTAACGGGGGGAATAGTCCGGC 120

Db 61 GCGGTGTTCCACATGCGGCGCGGTATATACCGCGCGCAATGTTCCGCGC 120
Qy 121 CCGACTACAGCTTGAAGATTTATCAGTACGTTTCCGCTTACGCTGCGTCTGCA 180
Db 121 CCGACTACAGCTTGAAGATTTATCAGTACGTTTCCGCTTACGCTGCGTCTGCA 180
Qy 181 AGCGATGCGGTAAATATGATCAGCTGTTACCGGTGTTTACCGATGAATGCGCAT 240
Db 181 CCGGTACCGGTGTTTACCGATGAATGCGCATGAGTTATGTAACGCGCGAT 240
Qy 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Db 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Qy 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Db 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Qy 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Qy 421 GGTTTGGCAACAAGCGCGCTTACCAATTTAA 456
Db 421 GGTTTGGCAACAAGCGCGCTTACCAATTTAA 456

RESULT 5

US-09-543-407-17
Sequence 17, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 17
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-17

Query Match
Best Local Similarity 79.6%; Score 363.2; DB 23; Length 456;
Matches 198; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 1 ATGAACTTTAAAGTGGCAGCTTGGCAGCATGTTGCTGCGAGTCTGCTGCT 60
Db 1 ATGAACTTTAAAGTGGCAGCTTGGCAGCATGTTGCTGCGAGTCTGCTGCT 60
Qy 61 GCGGTGTTCCACATGCGGCGCGGTATATCAATACGCGCGGCAATGTTCCGCGC 120
Db 61 GCGGTGTTCCACATGCGGCGCGGTATATCAATACGCGCGGCAATGTTCCGCGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACGCTGCTGCTGCA 180
Qy 181 AGCGATGCGGTAAATATGATCAGCTGTTACCGGTGTTTACCGATGAATGCGCAT 240
Db 181 AGCGATGCGGTAAATATGATCAGCTGTTACCGGTGTTTACCGATGAATGCGCAT 240

Qy 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Db 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Qy 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Db 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Qy 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Qy 421 GGTTTGGCAACAAGCGCGCTTACCAATTTAA 456
Db 421 GGTTTGGCAACAAGCGCGCTTACCAATTTAA 456

RESULT 6

US-09-543-407-11
Sequence 11, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant *Salmonella enteritidis* 3b afgA
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-11

Query Match
Best Local Similarity 79.3%; Score 361.6; DB 23; Length 456;
Matches 197; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 1 ATGAACTTTAAAGTGGCAGCTTGGCAGCATGTTGCTGCGAGTCTGCTGCT 60
Db 1 ATGAACTTTAAAGTGGCAGCTTGGCAGCATGTTGCTGCGAGTCTGCTGCT 60
Qy 61 GCGGTGTTCCACATGCGGCGCGGTATATCAATACGCGCGGCAATGTTCCGCGC 120
Db 61 GCGGTGTTCCACATGCGGCGCGGTATATCAATACGCGCGGCAATGTTCCGCGC 120
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACGCTGCTGCTGCA 180
Db 121 CCGGACTCAACGTTGAGCATTTATCAGTACGTTTCCGCTTACGCTGCTGCTGCA 180
Qy 181 AGCGATGCGGTAAATATGATCAGCTGTTACCGGTGTTTACCGATGAATGCGCAT 240
Db 181 AGCGATGCGGTAAATATGATCAGCTGTTACCGGTGTTTACCGATGAATGCGCAT 240
Qy 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Db 241 GCGAGCCAGGCGGTGATTAATAGTACTATTGAATGAGTCTGCAATGTTTCA 300
Qy 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Db 301 GCGACCATGACAGTGAAGCGCTAAACCTCGATTTACTGTGCGCATACGCGGT 360
Qy 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420
Db 361 AATAACGCGCGGTGTTATGATCAGACCGCATCTGATTCACCGTAAATGTTGCGTCA 420

Qy 421 GATTGGCAACAGCGCGCTAACGATATTA 456
Db 421 GCACATGCACACGCGCGCTAACGATATTA 456

RESULT 7

US-09-543-407-13
; Sequence 13, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 456
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
; OTHER INFORMATION: sequence containing the replacement fragment
; OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-13

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTCTGCTGCT 60
Qy 61 GCGCTGTTCCACATGCGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCG 120
Db 61 GCGCTGTTCCACATGCGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATGATGAGTTCGCTAACGCTGCTGCTGCTGCT 180
Db 121 CCGGACTCAACGTTGAGCATTTATGATGAGTTCGCTAACGCTGCTGCTGCTGCT 180
Qy 181 AGCGATGCGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 181 AGCGATGCGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Qy 241 GCAGCGCGAGGTGCGGTAAATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAGCGCGAGGTGCGGTAAATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GCGACATCGACGATGAGTGAAGCTAAACCTCCGATTAATCTGTCGCGCAATAG 360
Db 301 GCGACATCGACGATGAGTGAAGCTAAACCTCCGATTAATCTGTCGCGCAATAG 360
Qy 361 AATAAGCGCGGTGCTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 361 CTGGTACCGGTGCTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 421 GATTGGCAACAGCGCGCTAACGATATTA 456
Db 421 GATTGGCAACAGCGCGCTAACGATATTA 456

RESULT 8
US-09-543-407-23
; Sequence 23, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543.407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 23
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.
US-09-543-407-23

Query Match 79.3%; Score 361.6; DB 23; Length 456;
Best Local Similarity 87.1%; Pred. No. 8.2e-106;
Matches 397; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGCAGCATTCGACGATCGTAGTTCTGCGAGTCTGCTGCT 60
Qy 61 GCGCTGTTCCACATGCGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCG 120
Db 61 GCGCTGTTCCACATGCGGCGCGCGGTATCATTAACGGCGCGCAATAGTTCCGCG 120
Qy 121 CCGGACTCAACGTTGAGCATTTATGATGAGTTCGCTAACGCTGCTGCTGCTGCT 180
Db 121 CCGGACTCAACGTTGAGCATTTATGATGAGTTCGCTAACGCTGCTGCTGCTGCT 180
Qy 181 AGCGATGCGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 181 AGCGATGCGGTAAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Qy 241 GCAGCGCGAGGTGCGGTAAATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAGCGCGAGGTGCGGTAAATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 301 GCGACATCGACGATGAGTGAAGCTAAACCTCCGATTAATCTGTCGCGCAATAG 360
Db 301 GCGACATCGACGATGAGTGAAGCTAAACCTCCGATTAATCTGTCGCGCAATAG 360
Qy 361 AATAAGCGCGGTGCTGATTAATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 361 CATGAATGGCATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCA 420
Qy 421 GATTGGCAACAGCGCGCTAACGATATTA 456
Db 421 GATTGGCAACAGCGCGCTAACGATATTA 456

RESULT 9
US-09-543-407-15
; Sequence 15, Application US/09543407
; GENERAL INFORMATION:
; APPLICANT: White, Aaron P.
; APPLICANT: Doran, James L.
; APPLICANT: Collinson, S. Karen
; APPLICANT: Kay, William W.
; TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
; FILE REFERENCE: 920043.406
; CURRENT APPLICATION NUMBER: US/09/543.407
; CURRENT FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 59
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15

LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-15

Query Match
Best Local Similarity 78.9%; Score 360; DB 23; Length 456;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180
QY 181 AGGATGCGGCTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
DB 181 AGGATGCGGCTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
QY 241 GCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 300
DB 241 GCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 300
QY 301 GCGACCATGACAGTGAAGCTTAAACCTCCGATTTATCTGCTGCTGCTGCTGCT 360
DB 301 GCGACCATGACAGTGAAGCTTAAACCTCCGATTTATCTGCTGCTGCTGCTGCT 360
QY 361 AATAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 420
DB 361 AATAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 420
QY 421 GGTGTTGGCAACGCGCAGCTAATGCAATTTAA 456
DB 421 GGTGTTGGCAACGCGCAGCTAATGCAATTTAA 456

RESULT 10
US-09-543-407-27

Sequence 27, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 27
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-27

Query Match
Best Local Similarity 78.9%; Score 360; DB 23; Length 456;
Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Matches 396; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180
QY 181 AGGATGCGGCTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
DB 181 AGGATGCGGCTAATATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 240
QY 241 GCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 300
DB 241 GCAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 300
QY 301 GCGACCATGACAGTGAAGCTTAAACCTCCGATTTATCTGCTGCTGCTGCTGCT 360
DB 301 GCGACCATGACAGTGAAGCTTAAACCTCCGATTTATCTGCTGCTGCTGCTGCT 360
QY 361 AATAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 420
DB 361 AATAAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCA 420
QY 421 GGTGTTGGCAACGCGCAGCTAATGCAATTTAA 456
DB 421 GGTGTTGGCAACGCGCAGCTAATGCAATTTAA 456

RESULT 11
US-09-543-407-21

Sequence 21, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
TITLE OF INVENTION: BACTERIAL FIMBRIAL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 456
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga
OTHER INFORMATION: sequence containing the replacement fragment
US-09-543-407-21

Query Match
Best Local Similarity 78.6%; Score 358.4; DB 23; Length 456;
Matches 395; Conservative 0; Mismatches 61; Indels 0; Gaps 0;

QY 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
DB 1 ATGAACCTTTAAAGTGGAGCATTTGCGACATGCTAGTTCTGGAGTCTGGCT 60
QY 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
DB 61 GGGCTGCTTCCAAATGGGCGGCGGCTAATCATACGGGCGGCAATAGTTCCGGC 120
QY 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180
DB 121 CCGGACTCAAGTGGAGCATTTTACGATGCTGCTTACGCTGCTGCTGCTGCA 180

Db 121 CGGAGCTCAGGTTGAGCATTTATCAGTACGGTTCGGTAAACGCGCTGCTCTGCA 180
Qy 181 AGCGATGCGGCTTAATATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 240
Db 181 AGCGATGCGGCTTAATATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 240
Qy 241 GCGAGCCAGGGTGGGATTAATGACTATGTAAGTAACTGACCTCAAGATGTTTCAAGAAATAT 300
Db 241 GCGAGCCAGGGTGGGATTAATGACTATGTAAGTAACTGACCTCAAGATGTTTCAAGAAATAT 300
Qy 301 GCCACATGACGACGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCAATGACGACGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGCGCGT 360
Qy 361 AATAACGCGCGCTGTTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 420
Db 361 AATAACGCGCGCTGTTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 420
Qy 421 GGTTTGGCAACAACGCGCTTAACGATTTAA 456
Db 421 GGTTTGGCAACAACGCGCTTAACGATTTAA 456

RESULT 12

US-09-543-407-29

Sequence 29, Application US/09543407

GENERAL INFORMATION:

APPLICANT: White, Aaron P.

APPLICANT: Doran, James L.

APPLICANT: Collinson, S. Karen

APPLICANT: Kay, William W.

TITLE OF INVENTION: BACTERIAL FIBRILLAR SYSTEM FOR

FILE REFERENCE: 920043.406

CURRENT FILING DATE: 2000-04-05

NUMBER OF SEQ ID NOS: 59

SOFTWARE: PasteSeq for Windows Version 4.0

SEQ ID NO 29

LENGTH: 456

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Recombinant Salmonella enteritidis 3b afga

OTHER INFORMATION: sequence containing the replacement fragment

OTHER INFORMATION: encoding PT3 from GP63 of Leishmania major.

US-09-543-407-29

Query Match 77.9%; Score 355.2; DB 23; Length 456;

Best Local Similarity 86.2%; Pred. No. 9,7e-104; Indels 0; Gaps 0;

Matches 393; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

Qy 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTTCTGAGAGTCTGCTGCT 60
Db 1 ATGAACTTTTAAAGTGGAGCATTCGACCAATCGAGTTTCTGAGAGTCTGCTGCT 60
Qy 61 GCGCTGTTTCAACAATGGGCGCGCGGCTAATCAAAACGCGCGCAATAGTTCCGCG 120
Db 61 GCGCTGTTTCAACAATGGGCGCGCGGCTAATCAAAACGCGCGCAATAGTTCCGCG 120
Qy 121 CGGAGCTCAAGGTTGAGATTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 180
Db 121 CGGAGCTCAAGGTTGAGATTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 180
Qy 181 AGCGATGCGGCTTAATATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 240
Db 181 AGCGATGCGGCTTAATATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 240
Qy 241 GCGAGCCAGGGTGGGATTAATGACTATGTAAGTAACTGACCTCAAGATGTTTCAAGAAATAT 300
Db 241 GCGAGCCAGGGTGGGATTAATGACTATGTAAGTAACTGACCTCAAGATGTTTCAAGAAATAT 300

Qy 301 GCCACATGACGACGATGGAACGCTTAAAACTCCGATATTAATGTCGCGCAATACGCGCGT 360
Db 301 GCCACATGATGATGACTGCTGTTTAAACCGCTGTTTATCCCATGAAATGGCAATGCAAGCGCGT 360
Qy 361 AATAACGCGCGCTGTTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 420
Db 361 AATAACGCGCGCTGTTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 420
Qy 421 GGTTTGGCAACAACGCGCTTAACGATTTAA 456
Db 421 GGTTTGGCAACAACGCGCTTAACGATTTAA 456

RESULT 13

US-08-233-642A-54

Sequence 54, Application US/08233642A

GENERAL INFORMATION:

APPLICANT: Kay, William W.

APPLICANT: Collinson, S. Karen

APPLICANT: Clouthier, Sharon C.

APPLICANT: Doran, James L.

TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SALMONELLA-

NUMBER OF SEQUENCES: 58

NUMBER OF SEQUENCES: -

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: U.S.A.

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/233,642A

FILING DATE: 26-APR-1994

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: King, Joshua

REGISTRATION NUMBER: 35,570

REFERENCE/DOCKET NUMBER: 920043.403C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 682-6031

TELEFAX: (206) 682-6031

TELEX: 3723836 SEEDANBERRY

INFORMATION FOR SEQ ID NO: 54:

SEQUENCE CHARACTERISTICS:

LENGTH: 361 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

FEATURES:

NAME/KEY: CDS

LOCATION: 1..357

US-08-233-642A-54

Query Match 63.3%; Score 288.6; DB 6; Length 361;

Best Local Similarity 91.3%; Pred. No. 3.4e-82; Indels 0; Gaps 0;

Matches 306; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

Qy 64 GTCCTTCCACATAGGGGCGGCGGCTAATCAATACGCGCGCAATAGTTCCGCGCG 123
Db 1 GTCCTTCCACATAGGGGCGGCGGCTAATCAATACGCGCGCAATAGTTCCGCGCG 123
Qy 124 GACTCAACGTTGAGCATTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 183
Db 61 GACTCAACGTTGAGCATTTATGATCAGTGGTTACCGCTGTTTATCCCATGAAATGGCAAT 183

QY 184 GATCCCGTAATATGATAGCTGTTACCCGTTGTTATACCATGAATGCAATGCA 243
DB 121 GATGCCCGTAATATGTAACGACCAATACCCAGACGGGTTATGTAAGCGGCGGATGTA 180
QY 244 GGCACAGGTCGCGAATATAGTACTATGTAAGTCACTGCAATGCTTTCAGAAATATATGCC 303
DB 181 GGCACAGGTCGCGAATATAGTACTATGTAAGTCACTGCAATGCTTTCAGAAATATATGCC 240
QY 304 ACCATGACCACTGGAAGCGCTTAAATCTCGAATATTAATGTCGCGCAATACGGCGGTAAT 363
DB 241 ACCATGACCACTGGAAGCGCTTAAATCTCGAATATTAATGTCGCGCAATACGGCGGTAAT 300
QY 364 AACCGCGCGCTGCTTATCAGACCGCAATCTGATTC 398
DB 301 AACCGCGCGCTGCTTATCAGACCGCAATCTGATTC 335

RESULT 14

US-09-543-407-3
Sequence 3, Application US/09543407
GENERAL INFORMATION:
APPLICANT: White, Aaron P.
APPLICANT: Doran, James L.
APPLICANT: Collinson, S. Karen
APPLICANT: Kay, William W.
TITLE OF INVENTION: BACTERIAL FIBRIL SYSTEM FOR
FILE REFERENCE: 920043.406
CURRENT APPLICATION NUMBER: US/09/543,407
CURRENT FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 59
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 456
TYPE: DNA
ORGANISM: E. coli
US-09-543-407-3

Query Match 51.9%; Score 236.8; DB 23; Length 456;
Best Local Similarity 70.0%; Pred. No. 2,3e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGATTTTGGACGCTCTGGCT 60
DB 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGATTTTGGACGCTCTGGCT 60
QY 61 GCGCTGCTTCCAAATGGGCGCGGCTAATCATACGCGCGGCAATAGTTCCGCG 120
DB 61 GCGCTGCTTCCAAATGGGCGCGGCTAATCATACGCGCGGCAATAGTTCCGCG 120
QY 121 CCGAATCTGAGTGAACATTTACGATGAGGTCGCTTACCGCTGCTGCTGCA 180
DB 121 CCGAATCTGAGTGAACATTTACGATGAGGTCGCTTACCGCTGCTGCTGCA 180
QY 181 AGCGATCCCGTAATATGATCACTGCTGTTACCGGTTGTTACCGTAATAGGCAAT 240
DB 181 AGCGATCCCGTAATATGATCACTGCTGTTACCGGTTGTTACCGTAATAGGCAAT 240
QY 241 GCAAGCCAGGTCGCGAATATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 241 GCAAGCCAGGTCGCGAATATGATGATGATGATGATGATGATGATGATGATGAT 300
QY 301 GCGACATGCACTGAGTGAACGCTTAAATCTCGAATATTAATGTCGCGCAATAGGCGG 360
DB 301 GCGACATGCACTGAGTGAACGCTTAAATCTCGAATATTAATGTCGCGCAATAGGCGG 360
QY 361 AATAAGCCGCGCTGCTGTTATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 361 AATAAGCCGCGCTGCTGTTATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTGTTGCAACAAAGCGCAAGGCTTACCAATATTA 456
DB 421 GGTGTTGCAACAAAGCGCAAGGCTTACCAATATTA 456

RESULT 15

US-08-978-878-3
Sequence 3, Application US/08978878
GENERAL INFORMATION:
APPLICANT: OLSEN, Arne
APPLICANT: NORMARK, Staffan
TITLE OF INVENTION: FIBRONECTIN BINDING PROTEIN AS WELL AS ITS PREPARATION
FILE REFERENCE: 012889-081
CURRENT APPLICATION NUMBER: US/08/978,878
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: SE 8801723-1
EARLIER FILING DATE: 1988-05-06
EARLIER APPLICATION NUMBER: US 07/347,189
EARLIER FILING DATE: 1989-05-04
EARLIER APPLICATION NUMBER: US 07/789,437
EARLIER FILING DATE: 1991-11-06
EARLIER APPLICATION NUMBER: US 07/970,846
EARLIER FILING DATE: 1992-11-03
EARLIER APPLICATION NUMBER: US 08/187,865
EARLIER FILING DATE: 1994-01-28
EARLIER APPLICATION NUMBER: US 08/318,519
EARLIER FILING DATE: 1994-10-05
EARLIER APPLICATION NUMBER: US 08/495,959
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)...(538)
US-08-978-878-3

Query Match 51.9%; Score 236.8; DB 13; Length 675;
Best Local Similarity 70.0%; Pred. No. 2,6e-65;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

QY 1 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGATTTTGGACGCTCTGGCT 60
DB 83 ATGAAACTTTTAAAGTGGACGATTCGACGATTCGATTTTGGACGCTCTGGCT 60
QY 61 GCGCTGCTTCCAAATGGGCGCGGCTAATCATACGCGCGGCAATAGTTCCGCG 120
DB 143 GCGCTGCTTCCAAATGGGCGCGGCTAATCATACGCGCGGCAATAGTTCCGCG 120
QY 121 CCGAATCTGAGTGAACATTTACGATGAGGTCGCTTACCGCTGCTGCTGCA 180
DB 203 CCGAATCTGAGTGAACATTTACGATGAGGTCGCTTACCGCTGCTGCTGCA 180
QY 181 AGCGATCCCGTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
DB 263 AGCGATCCCGTAATATGATGATGATGATGATGATGATGATGATGATGATGAT 240
QY 241 GCAAGCCAGGTCGCGAATATGATGATGATGATGATGATGATGATGATGATGAT 300
DB 323 GTTGTCAGGCGCTGAGTGAACGCTTAAATCTCGAATATTAATGTCGCGCAATAGGCGG 382
QY 301 GCGACATGCACTGAGTGAACGCTTAAATCTCGAATATTAATGTCGCGCAATAGGCGG 360
DB 383 GCGACATGCACTGAGTGAACGCTTAAATCTCGAATATTAATGTCGCGCAATAGGCGG 360
QY 361 AATAAGCCGCGCTGCTGTTATGATGATGATGATGATGATGATGATGATGATGAT 420
DB 443 GCGAAGCGGTCGCGAATATGATGATGATGATGATGATGATGATGATGATGAT 420
QY 421 GGTGTTGCAACAAAGCGCAAGGCTTACCAATATTA 456
DB 503 GGTGTTGCAACAAAGCGCAAGGCTTACCAATATTA 456

Thu Mar 18 12:28:03 2004

us-09-543-407-19.rmpm

Page 9

Search completed: March 18, 2004, 02:47:09
Job time : 2925.24 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 16, 2004, 15:39:56 ; Search time 69.6301 Seconds
(without alignments)
2406.048 Million cell updates/sec

Title: US-09-543-407-19

Perfect score: 456

Sequence: 1 atgaacttctaaagtgsc.....ccagcgctacacagatratraa 456

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 637880 seqs, 183698769 residues

Total number of hits satisfying chosen parameters: 1275760

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents NA New :*

- 1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
- 2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
- 3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
- 4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
- 5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
- 6: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
- 7: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	236.8	51.9	675	5	US-09-741-873C-3
2	192.8	42.3	396	5	US-09-741-873C-1
3	30.8	6.8	1575	6	US-10-788-792-27
4	30.4	6.7	848	6	US-10-100-683-643
5	29.4	6.4	300	6	US-10-767-701-28437
6	29.2	6.4	249	6	US-10-771-241-117
7	29.2	6.4	757	6	US-10-771-241-29
8	29	6.4	1046	6	US-10-048-770C-3
9	28.8	6.3	3816	6	PCT-US04-04280-78
10	28.8	6.3	83009	6	US-10-417-375A-143
11	28.6	6.3	1186	1	PCT-US04-05654-1030
12	28.6	6.3	1295	6	US-10-100-683-5480
13	28.6	6.3	3416	6	US-10-100-683-3397
14	28.6	6.3	8206	6	US-10-021-698A-924
15	28.4	6.2	394468	7	US-60-548-091-5725
16	28.2	6.2	201	7	US-60-548-091-32688
17	28.2	6.2	596	6	US-10-767-701-5378
18	28.2	6.2	610	6	US-10-767-701-8781
19	28.2	6.2	1746	1	PCT-US04-05654-2555
20	28.2	6.2	14449	6	US-10-021-698A-924
21	28.2	6.2	225724	6	US-10-021-698A-711
22	28	6.1	237	6	US-10-767-701-31417
23	28	6.1	554	6	US-10-767-701-37293
24	27.8	6.1	749	6	US-10-767-701-12516
25	27.8	6.1	1400	7	US-60-545-213-2134
26	27.8	6.1	1400	7	US-60-545-213-2135

C	27	27.8	6.1	1400	7	US-60-545-213-6406	Sequence 6406, Ap
C	28	27.8	6.1	1400	7	US-60-545-213-6407	Sequence 6407, Ap
C	29	27.8	6.1	1422	6	US-10-767-795-3735	Sequence 3735, Ap
C	30	27.6	6.1	363	6	US-10-767-701-29572	Sequence 29572, A
C	31	27.6	6.1	463	6	US-10-767-701-17923	Sequence 17923, A
C	32	27.6	6.1	529	6	US-10-767-701-19620	Sequence 19620, A
C	33	27.6	6.1	551	6	US-10-767-701-3483	Sequence 3483, Ap
C	34	27.6	6.1	600	7	US-60-545-213-2020	Sequence 2020, Ap
C	35	27.6	6.1	607	7	US-60-545-213-6292	Sequence 6292, Ap
C	36	27.6	6.1	617	6	US-10-781-469-57	Sequence 57, Appl
C	37	27.6	6.1	632	6	US-10-100-683-3854	Sequence 3854, Ap
C	38	27.6	6.1	946	6	US-10-784-004-232	Sequence 232, Appl
C	39	27.6	6.1	1052	6	US-10-767-701-12607	Sequence 12607, A
C	40	27.6	6.1	1818	6	US-10-773-236-118	Sequence 118, Appl
C	41	27.6	6.1	4697	6	US-10-453-372-1185	Sequence 1185, Ap
C	42	27.6	6.1	6224	6	US-10-453-372-1173	Sequence 1173, Ap
C	43	27.6	6.1	6494	6	US-10-453-372-1187	Sequence 1187, Ap
C	44	27.6	6.1	8654	6	US-10-021-698A-3499	Sequence 3499, A
C	45	27.6	6.1	25426	6	US-10-100-683-11419	Sequence 11419, A

ALIGNMENTS

RESULT 1
US-09-741-873C-3
Sequence 3, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Oleon, Arne
TITLE OF INVENTION: Ribonectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT FILING DATE: US/09/741, 873C
PRIOR APPLICATION NUMBER: 2000-12-22
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978, 878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347, 189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789, 437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970, 846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187, 865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318, 519
PRIOR FILING DATE: 1994-10-05
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 675
TYPE: DNA
ORGANISM: Escherichia coli
FEATURE:
NAME/KEY: CDS
LOCATION: (83)..(538)
US-09-741-873C-3
Query Match 51.9%; Score 236.8; DB 5; Length 675;
Best Local Similarity 70.0%; Pred. No. 3.2e-72;
Matches 319; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 1 ATGAACCTTTAAAGTGGAGCATTCGAGCATCGTACTTTCGCGAGTCTGGCT 60
DB 83 ATGAACCTTTAAAGTGGAGCATTCGAGCATCGTACTTTCGCGAGTCTGGCA 142
QY 61 GGGGTGTTCCAAATGCGGCGCGCGGTAATCATTAACGCGCGCATATGTCGCG 120
DB 143 GGGGTGTTCTCAAGTACGCGCGCGGTAACCAACGCTGTCGCGTAATTAAGCGGC 202
QY 121 CCGACTCAACGTTGAGCATTTATCAATGCTTCGCTTAACGCTGCGCTTGCATGCA 180

Db 203 CCAATTCGTAGCTGATGACATTTACAGTACGGTGGCGTAACTCTGCACTTGTGCTGCA 262
Qy 181 AGCGATGCGGCTTAATATGATCAGCTGTTACCGGTTGTTATCCCATGAAATGACACAT 240
Db 263 ACTATATCCCGTAACTCTGACTTGAATATACCAGATGGCGGGTAAATGCTGACAT 322
Qy 241 GCAAGCGAGGTTGGGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 323 GTTGTGATGAGGCTCAGTACAGCTCAATGATGATGATGATGATGATGATGATGATGAT 382
Qy 301 GCCACCATGACAGTGAAGAGCTTAACTCCGATATTAAGTGTGCGGCAATACGGCGGT 360
Db 383 GCTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 442
Qy 361 AATAAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 443 GCGAAGCGGCTCAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 502
Qy 421 GGTGTTGGCAACAGCGGCAAGCTTAACTTAA 456
Db 503 GCGTTGTGTAACAGCGGCAAGCTTAACTTAA 538

RESULT 2

US-09-741-873C-1
Sequence 1, Application US/09741873C
GENERAL INFORMATION:
APPLICANT: Normark, Staffan
APPLICANT: Olsen, Arne
TITLE OF INVENTION: Fibronectin Binding Protein As Well As Its Preparation
FILE REFERENCE: 012889-084
CURRENT APPLICATION NUMBER: US/09/741,873C
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: SE 8801723-1
PRIOR FILING DATE: 1998-05-06
PRIOR APPLICATION NUMBER: US 08/978,878
PRIOR FILING DATE: 1997-11-26
PRIOR APPLICATION NUMBER: US 07/347,189
PRIOR FILING DATE: 1989-05-04
PRIOR APPLICATION NUMBER: US 07/789,437
PRIOR FILING DATE: 1991-11-06
PRIOR APPLICATION NUMBER: US 07/970,846
PRIOR FILING DATE: 1992-11-03
PRIOR APPLICATION NUMBER: US 08/187,865
PRIOR FILING DATE: 1994-01-28
PRIOR APPLICATION NUMBER: US 08/318,519
NUMBER OF SEQ ID NOS: 11
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 396
TYPE: DNA
ORGANISM: Escherichia coli
US-09-741-873C-1

Query Match 42.3%; Score 192.8; DB 5; Length 396;
Best Local Similarity 67.9%; Pred. No. 4,4e-57;
Matches 269; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

Qy 61 GGGGTGTTCCCAATGAGGCGGCGGCGGATCATATACGGCGGCAATAGTCCGCG 120
Db 1 GGTGTGTTCTCTAGTACGCGCGGCGGCGGTAACCAAGTGTGGGTAAATATAGCGCG 60
Qy 121 CCGGACTCAACGTTGAGCATTTATCAGTACGGTCCGTTCCGTTACGCTGCTCTGCA 180
Db 61 CCAATTTGAGTGAAGTGAACATTTACAGTACGGTGGGCGGTAAGTCTGCTCTGCA 120
Qy 181 AGCGATGCGGTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
Db 121 ACTGATGCGGTTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 180
Qy 241 GCAAGCGAGGTTGGGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Db 241 GCAAGCGAGGTTGGGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 300

Db 181 GTTGTGAGGCTCAGTACAGTACAGTCAATGATGATGATGATGATGATGATGATGAT 240
Qy 301 GCCACCATGACAGTGAAGAGCTTAACTCCGATATTAAGTGTGCGGCAATACGGCGGT 360
Db 241 GCTACTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 361 AATAAGCGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Db 301 GCGAAGCGGCTCAGTACGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 421 GGTGTTGGCAACAGCGGCAAGCTTAACTTAA 456
Db 361 GCGTTGTGTAACAGCGGCAAGCTTAACTTAA 396

RESULT 3

US-10-788-792-27/c
Sequence 27, Application US/10788792
GENERAL INFORMATION:
APPLICANT: Bayer Pharmaceuticals Corporation
APPLICANT: Eysenck, Deepa
APPLICANT: Bigwood, Douglas
TITLE OF INVENTION: EXPRESSION PROFILES FOR BREAST CANCER AND METHODS OF USE
FILE REFERENCE: 5152
CURRENT APPLICATION NUMBER: US/10/788,792
PRIOR FILING DATE: 2004-02-27
PRIOR APPLICATION NUMBER: US 60/450,655
PRIOR FILING DATE: 2003-02-28
NUMBER OF SEQ ID NOS: 254
SOFTWARE: PatentIn version 3.2
SEQ ID NO 27
LENGTH: 1575
TYPE: DNA
ORGANISM: Homo sapiens
US-10-788-792-27

Query Match 6.8%; Score 30.8; DB 6; Length 1575;
Best Local Similarity 55.7%; Pred. No. 1.1;
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 317 GGAAGCTTAAATCCGATTTACTGTGCGGCAATAGCGGCGGTAATAGCGCGGCTGG 376
Db 688 GGAATTCATACAGTCTGACCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 629
Qy 377 TTAATGAGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 422
Db 628 TGCACTGCTGCT 583

RESULT 4

US-10-100-683-643
Sequence 643, Application US/10100683
GENERAL INFORMATION:
APPLICANT: Rosen, et al.
TITLE OF INVENTION: Human Secreted Proteins
FILE REFERENCE: PS900
CURRENT APPLICATION NUMBER: US/10/100,683
PRIOR FILING DATE: 2002-03-19
PRIOR APPLICATION NUMBER: US 60/040,162
PRIOR FILING DATE: 1997-03-07
PRIOR APPLICATION NUMBER: US 60/043,576
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,601
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,845
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,580
PRIOR FILING DATE: 1997-04-11
PRIOR APPLICATION NUMBER: US 60/047,599
PRIOR FILING DATE: 1997-05-23
PRIOR APPLICATION NUMBER: US 60/056,664
PRIOR FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/043,314

Qy	255	AATATATG	CACATCGAC	TCGAC	CAAGTGAA	AGCTTAAAA	CTCCGAT	TTTCTGTCG	CGCCAA	TAAC	354
Db	264	AAATTTGA	AGCCATCGA	CTTATCA	CAAAAGCCAA	CAAGGTG	TGATCTCAT	CCGCGA	CTAC		323
Qy	355	GGCGGTAA	TAAAGCGCG	GTGTTAA	CAACA	CCGCAT	CTGATTC	CAAGC	CTAATG	TGCTGT	414
Db	324	GGCACC	CGACTT	CCCCC	ACAGAG	TAGTCAT	TAAAGAA	TTCACT	CTTC	TACTCG	383
Qy	415	CA	416								

RESULT 9.04-04280-78/C
PCT-US04-04280-78/C
Sequence 78, Application PC/US0404280
GENERAL INFORMATION:
APPLICANT: INCYTE CORPORATION; KABLE, Amy E.;
APPLICANT: YUE, Henry; BAUGHN, Mariah R.;
APPLICANT: TRIPOULEY, Catherine M.; RING, Huifun Z.;
APPLICANT: RANKIN, Uyen K.; EMBELING, Brooke E.;
APPLICANT: RAMAKRAN, Jayalaxmi; HARELIL, April J.A.;
APPLICANT: SAMANAKAR, Anita; LEE, Soo Yenui;
APPLICANT: CHANMLA, Narinder K.; GIERZEN, Kimberly J.;
APPLICANT: MARQUIS, Joseph P.; ELLIOTT, Vicki S.;
APPLICANT: BECHA, Shaanya D.; FAVERO, Kristin D.;
APPLICANT: WANG, Jonathan T.; NARIDU, Sangeta;
APPLICANT: HARKINS, Phillip R.; JIN, Pel;
APPLICANT: CHEN, David
TITLE OF INVENTION: ENZYMES
FILE REFERENCE: PF-1507 PCT
CURRENT APPLICATION NUMBER: PCT/US04/04280
CURRENT FILING DATE: 2004-02-12
PRIOR APPLICATION NUMBER: US 60/447,246
PRIOR FILING DATE: 2003-02-12
PRIOR APPLICATION NUMBER: US 60/449,087
PRIOR FILING DATE: 2003-02-21
PRIOR APPLICATION NUMBER: US 60/450,622
PRIOR FILING DATE: 2003-02-26
PRIOR APPLICATION NUMBER: US 60/456,704
PRIOR FILING DATE: 2003-03-21
PRIOR APPLICATION NUMBER: US 60/463,194
PRIOR FILING DATE: 2003-04-15
PRIOR APPLICATION NUMBER: US 60/466,358
PRIOR FILING DATE: 2003-05-09
PRIOR APPLICATION NUMBER: US 60/475,532

	Query Match	Best Local	Similarity	48.8%	Score	28.6;	DB	6;	Length	83009;
	Matches	78;	Conservative	0;	Mismatches	82;	Indels	0;	Gaps	0;
QY	281	AGAATGTTT	CAGAAATATATGCCACATGCACGAGTGAAGCCTTAAAACTCCATATT	TA	340					
Db	14056	AGAAATTTT	AAAAAGATGCATATATCTCTAAAAAGAACCTTACTAGTATTCCTTTATTC		1399					
QY	341	CTGTGGCCAT	TACGGCGGTATATACGCCGCGCTGGTATATCAGACCGCATCTGATCCA		400					
Db	13996	CTTTAACTAT	CACTTAAGTTAACTCTTCATGAATATACAGAGGCAACCTATTACTTCA		1393					
QY	401	GGGTATGTG	CGTCAGAGTTGGTTTGGCAAGCCAC		440					
Db	13936	GAATATGAC	CCCTCAGATGATTTACAGCAGATGCCAC		13897					

PCT-US04-05654-1030
; Sequence 1030, Application PC/TUS0405654
; GENERAL INFORMATION:
; APPLICANT: Sherman, Bradley K
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Haake, Jacqueline E
; APPLICANT: Creelman, Robert A
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc J
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddie, James
; APPLICANT: Dubell III, Arnold N
; APPLICANT: Pineda, Omeira
; APPLICANT: Repetti, Peter
; APPLICANT: Century, Karen
; APPLICANT: Guttersen, Neal
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Brown, Pierre E
; APPLICANT: Kumamoto, Roderick W
; APPLICANT: Pilgrim, Marsha L
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS
; FILE REFERENCE: MBI-0047 PCT
; CURRENT APPLICATION NUMBER: PCT/US04/05654
; CURRENT FILING DATE: 2004-03-03
; PRIOR APPLICATION NUMBER: 10/374,780
; PRIOR FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: 10/675,852
; PRIOR FILING DATE: 2003-09-30
; NUMBER OF SEQ ID NOS: 2950
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1030
; LENGTH: 1186
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Predicted polypeptide sequence is orthologous to G652
PCT-US04-05654-1030

Query Match 6.3%; Score 28.6; DB 1; Length 1186;
Best Local Similarity 59.0%; Pred. No. 5.7;
Matches 49; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 44 CTGCGAGTCTCTGCTGCGCTGCTGCTCCACATGCGCGCGCGGCTATATACGCG 103
DB 606 CAGGAGCTGCTCCAGCGCGCGCGCTACGCGGCTGCGCGCGCGGTGTGTGTGCG 665
QY 104 GCGGCAATAGTTCGCGCGCGGAC 126
DB 666 GCTGCTACAACTGCGCGCGGAC 688

RESULT 12
US-10-100-683-5480/c
; Sequence 5480, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599

PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5480
; LENGTH: 1295
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-100-683-5480

Query Match 6.3%; Score 28.6; DB 6; Length 1295;
Best Local Similarity 49.7%; Pred. No. 5.9;
Matches 73; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 242 CAGGCGAGGTCGCGATATAGTACTATGACTGACTCAGATGTTTCAATATATG 301
DB 767 CGGACCTGCTGCCCCCTTAACCTCAAGTCACTGCTGACGTCGCCAGGCACTAATG 708
QY 302 CCACCATGACCACTGAGCGTAAATCCGATATATGTCGCCAATACGCGGTA 361
DB 707 ACAGCATCAATCAAGCTCATCATATGTGACACCCAGGAGCAGCCGCGCAAGAGTGTG 648
QY 362 ATACGCGCGCGCTGCTTATATCAGACCG 388
DB 647 ATACGCGCTGCGGGAATGAGACGG 621

RESULT 13
US-10-100-683-3397/c
; Sequence 3397, Application US/10100683
; GENERAL INFORMATION:
; APPLICANT: Rosen, et al.
; TITLE OF INVENTION: Human Secreted Proteins
; FILE REFERENCE: PS900
; CURRENT APPLICATION NUMBER: US/10/100,683
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: US 60/040,162
; PRIOR FILING DATE: 1997-03-07
; PRIOR APPLICATION NUMBER: US 60/043,576
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,601
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,845
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,580
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,599
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,664
; PRIOR FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/043,314
; PRIOR FILING DATE: 1997-04-11
; PRIOR APPLICATION NUMBER: US 60/047,632
; PRIOR FILING DATE: 1997-05-23
; PRIOR APPLICATION NUMBER: US 60/056,892
; PRIOR FILING DATE: 1997-08-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 13468
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3397
; LENGTH: 3416
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-100-683-3397

Query Match	6.3%;	Score 28.6;	DB 6;	Length 3416;
Best Local Similarity	49.7%;	Pred. No. 9.5;		
Matches 73;	Conservative	0;	Mismatches	74; Indels 0

QY 242 CAGGCGAGGGTGGCGAATTAAGTACTATTGAATGTTTCAGAAATGTA 301
Db 761 CGAGCCCTGTGGCCCTTAACCTCAGAGTCAGCTGGTCGACGTGCCAGGGCAGTAATG 702
QY 302 CCACATCGACGAGTGGAAACGTAAATACTCCGATATTACTGTGGCCAAATACGGCGGTA 361
Db 701 ACAGCATCAATCAGCTCATCTCATGTGTGACCCAGCAGCAACCGGCCAGAAAGATGTG 642
QY 362 ATACGCCGCGCTGTTAATCAGACCG 388
Db 641 ATTAACGCCCTGGCGGAAATTGGAGACGG 615

```

RESULT 14.
US-10-021-698A--3529/c
; Sequence 3529, Application US/10021698A
; GENERAL INFORMATION:
; APPLICANT: KEITH, TIM
; APPLICANT: LITTLE, RANDALL
; APPLICANT: VAN EERDMECH, PAUL
; APPLICANT: DUPUIS, JOSEE
; APPLICANT: DEL MASTRO, RICHARD
; APPLICANT: SIMON, JASON
; APPLICANT: ALLEN, KRISTINA
; APPLICANT: PANDIT, SUNIL
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: RESPIRATORY DISEASES AND OBESITY
; FILE REFERENCE: 2976-4040US1
; CURRENT APPLICATION NUMBER: US/10/021,698A
; CURRENT FILING DATE: 2001-10-22
; PRIOR APPLICATION NUMBER: 60/211,749
; PRIOR FILING DATE: 2000-06-14
; NUMBER OF SEQ ID NOS: 6160
; SOFTWARE: PatentIn 2.1
; SEQ ID NO 3529
; LENGTH: 8206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (401)
; OTHER INFORMATION: a, t, c or g
US-10-021-698A-3529

```

Query Match	6.3%	Score 28.6;	DB 6;	Length 8206;
Best Local Similarity	49.7%;	Pred. No. 15;		
Matches	73;	Conservative	0;	Mismatches 74;
				Indels 0;
				Gaps 0;

QY	242	CAGGGCCAGGGGCGCGATTAATTAAGTACTATTGAACTGACTCAGATGTTTCAGAAATAATG	301
Db	4174	CGGACCTCGTGTGCCCCCTTAACCTCAGAGAGTACAGGTGGCTGACAGCTGCCAGGGCAGTAATCTG	4115
QY	302	CCACCATGAGACAGAGGAAACGTTAAAACTCCGATATTACTGTGGGCCAATATACGGCGGTA	361
Db	4114	ACAGACATCATCAGGTCTATCACTATGTGCACCCAGCAGCACCGGGCCAGAAAGATGTG	4055
QY	362	ATAACCGCGCGCTGTATTATCAGACCG	388
Db	4054	ATTAAAGCCCTTGGCGGAATTGGAGACGG	4028

RESULT 15
US-60-548-091-5725
; Sequence 5725, Application US/60548091
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: STROKE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001506

```

; CURRENT APPLICATION NUMBER: US/60/548,093
; CURRENT FILING DATE: 2004-02-27
; NUMBER OF SEQ ID NOS: 2443
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5725

```

US-60-548-091-5725

Query Match	6.2%	Score 28.4;	DB 7;	Length 394468;
Best Local Similarity	56.4%;	Pred. NO. 1.1e+02;		
Matches 53;	Conservative 0;	Mismatches 41;	Indels 0;	Gaps 0;

QY	260	ATAGTCTCTTGAACGACACTCAGAAATGGTTTCAGAAATATATGCAACCATGACCAAGTGA	319
Db	208307	ATATTCACACTCTAATGCCCAATATTTGTGCTTGCGAGATATTAACCACTATGAGTTGCTGA	208366
QY	320	AGGCTAAATACCCGATATATCTAGTCGGCCATAA	353
Db	208367	AGCTTATATCTCCAGGTCCTATTAATCCTCTTACAGA	208400

```
Search completed: March 17, 2004, 08:25:46
Job time : 70.6301 secs
```